An introduction to stochastic partial differential equations via diffusion approximation

Julien Vovelle

An introduction to stochastic partial differential equations via diffusion approximation

Julien Vovelle

September 9, 2018

Contents

1 Introduction

2 An introduction to probability theory

2.1 Some examples

The object of probability theory is to investigate the phenomena which resort to chance. We will study random dynamical systems in that course: for such systems (see Examples 1 and 2 below), the intuition is that understanding the state at time n is more and more difficult as n increases. Probability theory will show, in part, that it is not the case. Random experiments, and possibly their repetitions, are not an indistinct chaos. The central notion in a scientific approach to random phenomena will be the notion of *law*: prescribing the law of chance of random events (or random variables) does not limit their uncertainty, but bound them to a well-defined mathematical object. See Section 2.3. The notion of *independence* is also prominent in probability theory (see Example 3 below) and will be introduced in Section 2.5.

Example 2.1 (Coin tossing). Consider the repeated tossing of an unbiased coin. When N of these experiments have been realized, what can we say? Certainly, what we do not know is the specific value of the list $R_1, \dots, R_N$ of the results, from step 1 to step N (R_i is head or tail according to the result obtained at the i-th tossing). We know however, among several possible things, and intuitively perhaps, that

1. the probability that $(R_1, \dots, R_N)$ is equal to a given element of $\{\text{head}, \text{tail}\}^N$ is 2^{-N},
2. the result of the i-th tossing is independent of the result of the j-th tossing ($1 \leq i < j \leq N$),
3. for large N the number of outcome of head should be similar to the number of outcome of tail, and thus similar to $\frac{N}{2}$.

Example 2.2 (Random walk). Consider a random walk on $\mathbb{Z}$, a process described by the following iteration: given the position $X_N \in \mathbb{Z}$ reached at time N, draw, independently on $X_0, \dots, X_N$, a random variable Z_{N+1} taking the values $+1$ or -1 with equi-probability and set

$$X_{N+1} = X_N + Z_{N+1}.$$

The initial position is taken at the origin in general: $X_0 = 0$. Once again, we do not know the position at time N or the trajectory $(X_0, \dots, X_N)$ up to time N. However, here is a list of some facts we know:

1. (for symmetry reasons) the probability that $X_N > 0$ is equal to the probability that $X_N < 0$,
2. *on average*, the position of X_N is zero,
3. X_N has the same parity as N,
4. *on average*, $|X_N|^2 = N$.

Here are some remarks on these results, and some arguments for the point 4. Note first that 3. is a deterministic statement: this is true whatever the outcomes may be. The other statements have a probabilistic nature (we speak of the probability of an event, or of some results on average). Item 2. is a consequence of 1. The arguments for the point 4. are the following ones: by developing the square, we have

$$|X_{N+1}|^2 = |X_N|^2 + 2X_N Z_{N+1} + |Z_{N+1}|^2 = |X_N|^2 + 2X_N Z_{N+1} + 1. \tag{2.1}$$

Indeed, $|Z_{N+1}|$ always takes the value 1. On average, $X_N Z_{N+1} = 0$ and thus, on average, $|X_{N+1}|^2 = |X_N|^2 + 1$, which gives the result by iteration. The rigorous version of this computation will be given once the notion of independence and expectancy have been introduced, see 2.29. As a final remark, consider the following question: what is, on average, the distance of X_N to its starting point after N steps? Item 4. gives the upper bound $\sqrt{N}$ since $\mathbb{E}|X_N| \leq [\mathbb{E}|X_N|^2]^{1/2}$ by the Cauchy-Schwarz Inequality. The notation $\mathbb{E}$ for the expectancy operator is introduced in Section 2.7.

Example 2.3 (Numbers). Draw two numbers a and b in $[0,1]$. What is the probability that $a \geq b$? The answer one-half comes to mind since the probability that $a \geq b$ seems to equal the probability that $b \geq a$. To be true, however, this explanation requires a and b to be drawn according to the same (continuous) *law*, and to be *independent* on each other. Here we meet again these two notions, that we will introduce in Section 2.3 and Section 2.5 respectively.

2.2 Probability space, random variable

Definition 2.1 (Probability space). A *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ is a measure space where the measure has total mass 1: $\mathcal{F}$ is a σ-algebra on Ω, a non-empty set, and $\mathbb{P}$ a measure on $(\Omega, \mathcal{F})$ such that $\mathbb{P}(\Omega) = 1$. This space is said to be *complete* when $\mathcal{F}$ contains all the negligible sets (definition: a subset A of Ω is said to be *negligible* if it is contained in a set $\tilde{A} \in \mathcal{F}$ such that $\mathbb{P}(\tilde{A}) = 0$). The elements of $\mathcal{F}$ are called *events*.

Exercise 2.2. Find the experiment corresponding to each of the following probability spaces and give the characterization of the event A in terms of this experiment.

1. $\Omega = \{1, \dots, 6\}$, $\mathbb{P}(\{i\}) = \frac{1}{6}$, $A = \{2, 4, 6\}$.
2. $\Omega = \{H, T\}^2$, $\mathbb{P}(\{\omega\}) = \frac{1}{4}$ for each $\omega \in \Omega$, $A = \{(H, T), (H, H)\}$.
3. $\Omega = \{\gamma \in C([0, T]; \mathbb{R}^2); \gamma(0) = 0\}$, $\mathbb{P} =$"to be seen later",

$$A = \{\gamma \in \Omega; \exists t \in [0, T], \gamma(t) \in D\},$$

where D is a closed subset of $\mathbb{R}^2$ (*e.g.* D is the closed disk of radius 1 and center $(2, 0)$).

Note: what may be $\mathcal{F}$ in the last example?

The solution to Exercise 2.2 is here.

Definition 2.3 (Random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable space. A map $X : \Omega \to E$ is said to be a *random variable on E* if it is measurable: for all $B \in \mathcal{E}, X^{-1}(B) \in \mathcal{F}$.

It is actually in terms of a random variable that the outcomes of random experiments are expressed.

Exercise 2.4. Come back to Exercise 2.2. In each case, introduce a natural random variable X and write the event A in terms of X.

The solution to Exercise 2.4 is here.

In general, E is a topological space and $\mathcal{E}$ the σ-algebra of the Borelians. All the events characterized by a random variable X form the following sub-σ-algebra of $\mathcal{F}$:

$$\sigma(X) = \{X^{-1}(B); B \in \mathcal{E}\}.$$

If $\Phi : (E, \mathcal{E}) \to (\tilde{E}, \tilde{\mathcal{E}})$ is a measurable application between two measurable spaces, then $Y := \Phi \circ X$ is $\sigma(X)$-measurable. Indeed, for all $\tilde{B} \in \tilde{\mathcal{E}}$, we have

$$Y^{-1}(\tilde{B}) = X^{-1}(\Phi^{-1}(\tilde{B})) \in \sigma(X).$$

Conversely, we have the following result.

Theorem 2.1. *Let E and $\tilde{E}$ be two separable Banach spaces endowed respectively with the σ-algebra $\mathcal{E}$ and $\tilde{\mathcal{E}}$ of Borelians. Let*

$$X : (\Omega, \mathcal{F}) \to (E, \mathcal{E}), \quad Y : (\Omega, \mathcal{F}) \to (\tilde{E}, \tilde{\mathcal{E}})$$

be two random variables. If Y is $\sigma(X)$-measurable, then there exists $\Phi : (E, \mathcal{E}) \to (\tilde{E}, \tilde{\mathcal{E}})$ measurable such that $Y = \Phi \circ X$.

To prove Theorem 2.1, we will use the following result of approximation by simple functions.

Definition 2.5 (Simple function)**.** Let E be a separable Banach space endowed with the Borel σ-algebra $\mathcal{E}$. A random variable $X\colon\Omega\to E$ is said to be *simple* if, almost-surely, it takes a finite number of values. Equivalently, X if simple if it can be written as

$$X=\sum_{i\in I}x_i\mathbf{1}_{A_i},$$

where I is a finite set, $x_i\in E$, $A_i\in\mathcal{F}$ and $\mathbf{1}_A$ is the characteristic function of the set A.

In Definition 2.5, we have used for the first time the term "almost-surely". Here is the definition

Definition 2.6 (Almost-sure)**.** An event $A\in\mathcal{F}$ is said to be *almost-sure*, or to be realized *almost-surely*, if $\mathbb{P}(A)=1$.

The result of approximation by simple functions can now be stated as follows.

Proposition 2.2 (Approximation by simple functions)**.** *Let E be a separable Banach space endowed with the Borel σ-algebra $\mathcal{E}$. If $X\colon\Omega\to E$ is a random variable, then there exists a sequence of simple functions (X_n) which converges almost-surely to X and such that $\|X_n\|_E\leq 2\|X\|_E$.*

Proof of Proposition 2.2. Let $E_\infty=\{x_k;k\in\mathbb{N}\}$ be a dense countable subset of E. We assume $x_0=0$. The random variables X takes values in the adherence of E_∞, which is E, hence X is not far from taking its values in E_∞, which itself is not far from being finite. To construct the sequence (X_n) properly, set $E_n=\{x_k;0\leq k\leq n\}$ and define the projection $p_n\colon E\to E_n$ by associating to $x\in E$ the closest element $y(x)$ of E_n. Such a $y(x)$ is well defined if

$$d(x,E_n)=\min\{\|x-y\|_E;y\in E_n\}$$

is realized for a single $y\in E_n$. If there are several points $y\in E_n$ for which the minimum is reached, we define $p_n(x)$ as the point x_k with lower index $k\in\{0,\dots,n\}$. The set $p_n^{-1}(\{x_k\})$ is therefore

$$\left[\bigcap_{l=k}^{n}\{x\in E;\|x-x_k\|_E\leq\|x-x_l\|_E\}\right]\bigcap\left[\bigcap_{l=0}^{k-1}\{x\in E;\|x-x_k\|_E<\|x-x_l\|_E\}\right].$$

In particular, the projection p_n is measurable and $X_n:=p_n\circ X$ is a simple function. Let us prove that (X_n) converges almost-surely to X. Actually, $X_n(\omega)\to X(\omega)$ for all $\omega\in\Omega$. Indeed, given $\varepsilon>0$, there exists $n\geq 0$ such that $\|X(\omega)-x_n\|_E<\varepsilon$. Then $\|X_m(\omega)-X(\omega)\|_E<\varepsilon$ for all $m\geq n$ by construction. Note that $\|x-p_n(x)\|_E\leq\|x\|$ since $0=x_0\in E_n$. By the triangular inequality, we deduce $\|p_n(x)\|_E\leq 2\|x\|_E$: this gives the bound $\|X_n\|_E\leq 2\|X\|_E$. $\square$

Proof of Theorem 2.1. Assume first that Y is simple, then

$$Y = \sum_{i \in I} y_i \mathbf{1}_{A_i},$$

where I is a finite set, $y_i \in \tilde{E}$ and $A_i \in \sigma(X)$. By definition, $A_i = X^{-1}(B_i)$, where $B_i \in \mathcal{E}$. Define $\Phi = \sum_{i \in I} y_i \mathbf{1}_{B_i}$. Then $\Phi\colon E \to \tilde{E}$ is measurable, and $Y = \Phi \circ X$. In the general case, consider a sequence (Y_n) of simple random variables which converges to Y. We apply Proposition 2.2 with the σ-algebra $\mathcal{F} = \sigma(X)$. Then each Y_n is $\sigma(X)$-measurable and, for each n, there exists $\Phi_n\colon E \to \tilde{E}$ measurable such that $Y_n = \Phi_n \circ X$. Introduce the Borel set B of the points of convergence of the sequence (Φ_n). We can use a Cauchy criterion to characterize B and show that it is indeed a Borel set:

$$B = \bigcap_{k \geq 1} \bigcup_{n \geq 1} \bigcap_{p,q \geq n} \left\{ x \in E; \|\Phi_p(x) - \Phi_q(x)\|_{\tilde{E}} < k^{-1} \right\}.$$

Define $\Phi(x) = \lim_{n \to +\infty} \Phi_n(x)$ if $x \in B$, $\Phi(x) = 0$ otherwise. Then $\Phi\colon E \to \tilde{E}$ is measurable and $Y = \Phi \circ X$. Indeed, for all $\omega \in \Omega$, $X(\omega) \in B$ since $\Phi_n(X(\omega)) = Y_n(\omega)$ converges to $Y(\omega)$. This concludes the proof. $\square$

2.3 The law of a random variable – I

Definition 2.7. Let E be a separable Banach space endowed with the Borel σ-algebra $\mathcal{E}$. Let $X\colon \Omega \to E$ be a random variable. The *law* of X is the measure μ_X on $(E, \mathcal{E})$ defined by

$$\mu_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B), \tag{2.2}$$

for all $B \in \mathcal{E}$.

Note the use of the probabilistic notation $\mathbb{P}(X \in B)$ in (2.2). In measure theory, μ_X is the image measure of $\mathbb{P}$ by X, or push-forward of $\mathbb{P}$ by X (notation $X_*\mathbb{P}$ or $X_\sharp\mathbb{P}$). This is a probability measure on $(E, \mathcal{E})$. Here are some examples of laws.

2.3.1 Bernoulli's law.

This is the law of a random variable X taking values in a set $E = \{x_1, x_2\}$ with two elements, according to the probabilities

$$\mathbb{P}(X = x_1) = p, \quad \mathbb{P}(X = x_2) = 1 - p, \tag{2.3}$$

where $p \in [0,1]$ is given. Otherwise speaking, this is the measure

$$p\delta_{x_1} + (1-p)\delta_{x_2}.$$

Note that x_1 and x_2 are generic here. The generic notation of the Bernoulli's law is $b(p)$. We write $X \sim b(p)$ if X satisfies (2.3) for some set $\{x_1, x_2\}$.

2.3.2 Exponential law.

This is the law of a random variable X taking values in $[0, +\infty)$ with probabilities

$$\mathbb{P}(X \geq t) = e^{-\lambda t}, \tag{2.4}$$

for all $t \geq 0$, where λ is a given positive parameter. Since

$$\mathbb{P}(X \geq t) = e^{-\lambda t} = \int_t^{+\infty} \lambda e^{-\lambda s} ds,$$

and since the sets $[t, +\infty)$ for $t \geq 0$ form a π-system which generates the Borel σ-algebra of $[0, +\infty)$, we see (*cf.* [Bil95, Theorem 3.3]) that the exponential law is the measure of density $t \mapsto \lambda e^{-\lambda t}$ with respect to the Lebesgue measure on $[0, +\infty)$. It is denoted by $\mathcal{E}(\lambda)$.

2.3.3 Binomial law.

This is the law, denoted $\mathcal{B}(n, p)$, of a random variable X taking values in $E = \{0, \ldots, n\}$ according to the probabilities

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \tag{2.5}$$

where $n \in \mathbb{N}^*$ and $p \in [0, 1]$.

Remark 2.4 (Vocabulary). Instead of *law*, one also speaks of *distribution* (Bernoulli's distribution, exponential distribution, and, see below, binomial distribution, Poisson distribution, normal distribution, etc.) Indeed, knowing the law of X is knowing how are distributed the possible values of X. The following exercise (which requires the notion of independence) is an illustration of this notion.

Exercise 2.8 (Invariant measure). Let $X_0, X_1, \ldots$ be the sequence of random variables on $\mathbb{R}$ defined as follows: X_0 is chosen at random, according to a law μ_0, then, X_N being known, a random variable Z_{N+1} taking the values $+1$ or -1 with equi-probability is drawn independently on $X_0, \ldots, X_N$ and X_{N+1} given by

$$X_{N+1} = \frac{1}{2} X_N + Z_{N+1}.$$

1. What means $\mu_0 = \delta_0$? What are then the law μ_1, μ_2 of X_1 and X_2 respectively?
2. Consider the case $\mu_0 = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_{+2}$. Compute μ_1, μ_2, μ_3. Can you guess a general formula for μ_N?
3. Find the way to choose μ_0 which ensures that μ_N, the law of X_N, is equal to μ_0 for all $N \geq 0$.

The solution to Exercise 2.8 is here.

Remark 2.5 (Important). Consider the statement "Let X be a random variable of law $b(p)$" or "Let Y be a random variable of law $\mathcal{E}(\lambda)$". In such statements, the nature of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is not specified. Is it problematic? Actually not, since the specification of the probability space is not relevant (illustration in the following exercise) and since the existence of a probability space and of such a random variable is ensured. To justify this last assertion, note that, if μ is a probability measure on a measure space $(E, \mathcal{E})$, and if

$$(\Omega, \mathcal{F}, \mathbb{P}) = (E, \mathcal{E}, \mu), \tag{2.6}$$

then the identity on E is a random variable of law μ.

Exercise 2.9. Let $X : \Omega \to E$ be a random variable. Show that, from the knowledge of the law μ_X of X follows the knowledge of all the probabilities $\mathbb{P}(A)$ for $A \in \sigma(X)$.

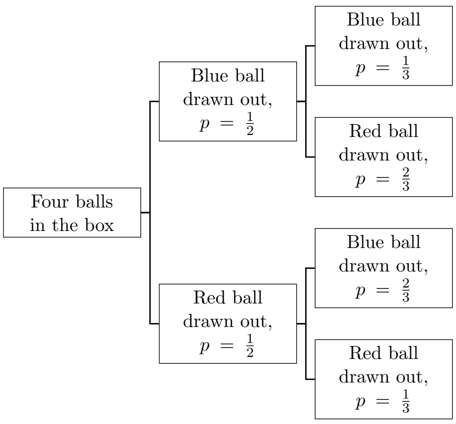
The solution to Exercise 2.9 is here.

We will give the relation between the Binomial law $\mathcal{B}(n, p)$ and the Bernoulli's law $b(p)$. We will also introduce some other laws (Poisson, Normal). To do this in a consistent way, we need to introduce first the notions of conditional probability and independence.

2.4 Conditional probability

An example. Two balls are drawn out successively from a box containing four balls initially: two red balls and two blue balls. What is the probability that the two balls drawn out from the box have the same color?

There are many ways to answer to this question (at least one combinatoric way and one probabilistic way). One can draw the following tree to conclude that the probability that the two balls have the same color is $\frac{1}{3}$.



This corresponds to the following equalities, where A is the event "the two balls have the same color" and B is the event "the first ball drawn out has the color red":

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) \\ &= \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2} = \frac{1}{3}.\end{aligned} \tag{2.7}$$

In (2.7), $\mathbb{P}(A|B)$ is the "probability of A knowing B", or, more precisely, the "probability of A knowing that B has been realized".

Definition 2.10 (Conditional probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let A, B be two events with $\mathbb{P}(B) > 0$. The probability of A conditionally to B is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{2.8}$$

Equation (2.7) is an instance of the formula of total probability.

Exercise 2.11. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Prove the following formula of total probability: if $A_1, \dots, A_n$ are disjoint events whose union has probability one and A an event, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|A_i)\mathbb{P}(A_i) \tag{2.9}$$

The solution to Exercise 2.11 is here.

2.5 Independence

Definition 2.12 (Independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two events A and $B \in \mathcal{F}$ are said to be *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{2.10}$$

Equivalently to (2.10), if $\mathbb{P}(B) > 0$, one has $\mathbb{P}(A|B) = \mathbb{P}(A)$: the knowledge of B has no influence on the realization of A. To test the definition of independence through (2.10), consider the basic example of one card drawn from a pack of 52 cards and the events $A =$"this is an ace", $B =$"this is a heart". We have the respective probabilities

$$\mathbb{P}(A) = \frac{4}{52} = \frac{1}{13}, \quad \mathbb{P}(B) = \frac{13}{52} = \frac{1}{4}, \quad \mathbb{P}(A \cap B) = \frac{1}{52} = \mathbb{P}(A)\mathbb{P}(B).$$

The events A and B are independent.

Beware of intuition in matter of independence. Consider for example the following experiment: one rolls two dices. The respective results are denoted X_1 and X_2. Consider the events

$$A_1 = \{X_1 + X_2 = 6\}, \quad A_2\{X_1 + X_2 = 7\}, \quad B = \{X_1 = 4\}.$$

Then A_1, the event that the sum of the dices is 6, and B, the event that the first dice gives 4, are *not* independent (as expected intuitively), but A_2 and B *are* independent.

Exercise 2.13. Justify the assertion above.

The solution to Exercise 2.13 is here.

The definition of independence for several events $A_1, \dots, A_n$ and for random variables $X_1, \dots, X_m$ is based on (2.10).

Definition 2.14 (Independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

1. The events $\{A_i; i \in I\}$ are said to be *independent* if, for all finite subset $J \subset I$, one has
$$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i).$$

2. The sub σ-algebras $\mathcal{F}_i \subset \mathcal{F}$ for $i \in I$ are said to be *independent* if, for all $A_i \in \mathcal{F}_i$, the $\{A_i; i \in I\}$ are independent.

3. The random variables $\{X_i; i \in I\}$ are *independent* if the σ-algebra $\sigma(X_i)$ for $i \in I$ are independent.

The following exercise illustrates a situation where this is not the independence of two random variables, but on the contrary a particular dependence between them (this is called a coupling), which is sought. The solution uses independence though.

Exercise 2.15 (Maximal coupling)**.** Let X and Y be two random variables uniformly distributed on $[0, 1]$ and $[0, 1/2]$ respectively:
$$\mathbb{P}(X \in A) = |A \cap [0, 1]|, \quad \mathbb{P}(Y \in A) = 2|A \cap [0, 1/2]|,$$
for all Borel subset A of $\mathbb{R}$, where $|A|$ is the Lebesgue measure of A. Find a way to draw X and Y maximizing the probability that $X = Y$, *i.e.* explain how to draw two random variables $\hat{X}$ and $\hat{Y}$ having same laws as X and Y respectively, which maximize $\mathbb{P}(\hat{X} = \hat{Y})$ among such random variables.

The solution to Exercise 2.15 is here.

This second exercise uses the notion of independence, and the probabilistic framework, to build an example of a sequence converging in L^1 but not a.e.

Exercise 2.16. Let (X_n) be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of respective law $b(\frac{1}{n})$:
$$\mathbb{P}(X_n = 1) = \frac{1}{n}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$
Show that $X_n \to 0$ in $L^1(\Omega, \mathbb{P})$ and that (X_n) does not converge to zero almost-surely.

The solution to Exercise 2.16 is here.

A fundamental example of independence is the case of a random variable X on $\mathbb{R}^n$ whose coordinates $X_1, \dots, X_n$ constitute independent random variables on $\mathbb{R}$. The law of X is then the product (in the sense of measures) of the laws of the coordinates.

Theorem 2.3. *Let $X_1, \dots, X_n$ be* independent *random variables on $\mathbb{R}$. Then the law of the $\mathbb{R}^n$-valued random variable $(X_1, \dots, X_n)$ is the product of the laws:*

$$\mu_{(X_1,\dots,X_n)} = \mu_{X_1} \times \cdots \times \mu_{X_n}. \tag{2.11}$$

Proof of Theorem 2.3. Let $X = (X_1, \dots, X_n)$. Let $\mathcal{R}_n$ denote the class of measurable rectangles: this is the class of all Borel subsets of $\mathbb{R}^n$ of the form

$$A = A_1 \times \cdots \times A_n,$$

where $A_1, \dots, A_n$ are Borel subset of $\mathbb{R}$. By independence, we have

$$\mathbb{P}(X_1 \in A_1 \ \& \ X_2 \in A_2 \ \& \ \dots \ \& \ X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n). \tag{2.12}$$

The left-hand side of (2.12) is $\mathbb{P}(X \in A)$, the right-hand side is $\mu_{X_1} \times \cdots \times \mu_{X_n}(A)$. Since $\mathcal{R}_n$ is a π-system, the probability measures μ_X and $\mu_{X_1} \times \cdots \times \mu_{X_n}$ coincide on the σ-algebra generated by $\mathcal{R}_n$, [Bil95, Theorem 3.3]. This latter is the whole class of the Borelians on $\mathbb{R}^n$ [Bil95, Example 18.1]. $\square$

Exercise 2.17. Generalize Theorem 2.3 to the case of Banach-valued random variables. Let $E_1, \dots, E_n$ be some separable Banach spaces. Let $X_1, \dots, X_n$ be *independent* random variables, X_i being a random variable on E_i. Then the law of the $E_1 \times \cdots E_n$-valued random variable $(X_1, \dots, X_n)$ is the product of the laws: (2.11) is satisfied.

The solution to Exercise 2.17 is here.

An important application of Theorem 2.3 is the computation of the law of the sum of two independent random variables. We give the statement for two random variables, but it can be generalized to any finite number of independent random variables by iteration.

Theorem 2.4 (Sum of independent random variables)**.** *Let X, Y be two* independent *random variables on $\mathbb{R}$. Then the law of $X + Y$ is the convolution product $\mu_X * \mu_Y$. In particular, if μ_X has density f_X with respect to the Lebesgue measure on $\mathbb{R}$ and μ_Y has density f_Y with respect to the Lebesgue measure on $\mathbb{R}$, then μ_{X+Y} is the measure of density*

$$f_{X+Y} = f_X * f_Y$$

with respect to the Lebesgue measure on $\mathbb{R}$.

Recall that the convolution product of two integrable functions $f, g \in L^1(\mathbb{R})$ is defined by

$$f * g(x) = \int_{\mathbb{R}} f(y) g(x - y) dy.$$

In particular, if $h \in C_b(\mathbb{R})$ (continuous bounded function), then, using Fubini's Theorem and a change of variable, we have

$$\int_{\mathbb{R}} h(x) f * g(x) dx = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x + y) f(x) dx g(y) dy.$$

This formula is then generalized into a definition of the convolution product of two Borel finite measures μ and ν:

$$\int_{\mathbb{R}} h(x) d\mu * \nu(x) := \int_{\mathbb{R}} \int_{\mathbb{R}} h(x+y) d\mu(x) d\nu(y), \tag{2.13}$$

for all $h \in C_b(\mathbb{R})$. If we introduce the function sum $\sigma \colon \mathbb{R}^2 \to \mathbb{R}$ defined by $\sigma(x,y) = x+y$, then (2.13) takes the more concise form

$$\mu * \nu = \sigma_\sharp(\mu \times \nu). \tag{2.14}$$

The convolution $\mu * \nu$ is the push-forward of the product measure $\mu \times \nu$ by σ.

Proof of Theorem 2.4. The law of $X+Y$ is the push-forward of $\mu_{(X,Y)}$ by σ. Indeed, if A is a Borel subset of $\mathbb{R}$, then $\mathbb{P}(X+Y \in A)$ is equal to

$$\mathbb{P}((X,Y) \in \sigma^{-1}(A)) = \mu_{(X,Y)}(\sigma^{-1}(A)) = \sigma_\sharp \mu_{(X,Y)}(A).$$

By independence and (2.11) and Formula (2.14) for the convolution product, we obtain the result. □

Let us illustrate the application of Theorem 2.4 by two examples. Consider first some independent variables $X_1, \ldots, X_n$, of Bernoulli's law $b(p)$:

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1-p.$$

The law of X_i is $\mu = p\delta_1 + (1-p)\delta_0$. The law of $X_1 + X_2$ is given by

$$\begin{aligned} \int_{\mathbb{R}} h d\mu_{X_1+X_2} &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x+y) d\mu(x) d\mu(y) \\ &= \int_{\mathbb{R}} ph(1+y) + (1-p)h(y) d\mu(y) \\ &= p^2 h(2) + 2p(1-p)h(1) + (1-p)^2 h(0). \end{aligned}$$

Two sum up, $\mu_{X_1+X_2} = p^2\delta_2 + 2p(1-p)\delta_1 + (1-p)^2\delta_0$. This is a Binomial law $\mathcal{B}(2,p)$. The generalization to n terms is given as an exercise.

Exercise 2.18. Let $X_1, \ldots, X_n$ be some independent variables of Bernoulli's law $b(p)$:

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1-p.$$

Show that $X_1 + \cdots + X_n$ follows the Binomial law $\mathcal{B}(n,p)$.

The solution to Exercise 2.18 is here.

2.6 The law of a random variable – II

2.6.1 Binomial law.

We come back to the Binomial law, introduced in paragraph 2.3.3. Consider the following experiment: repeat n times, successively and in an independent manner, a trial where each outcome has probability p of success, $1-p$ of failure. Such an experiment is called a *Bernoulli's test*. To such a test, we associate the following question: for $k \in \{0, \dots, n\}$, what is the probability to get k success precisely? The answer is the following one: define X_i for $1 \leq i \leq n$, by $X_i = 1$ if success occurs, $X_i = 0$ otherwise. Then

$$X := X_1 + \cdots + X_n$$

is the total number of success after n repetition of the experiment. The random variable $\{X_i; 1 \leq i \leq n\}$ are independent Bernoulli's $b(p)$. By Exercise 2.18, X follows a Binomial law $\mathcal{B}(n,p)$.

A classical example where Binomial law applies is the following one: consider a production's line in a factory. Each object released at the end of the line has a probability 0.01 to have a defect. In a set of 100 objects released, what is the probability to find at least one object which has a defect? The answer is certainly not $100 \times 0,01 = 1$ (what if the question was about 1,000 objects?). The answer is $\mathbb{P}(X \geq 1)$, were, assuming independence in the production of the objects, X is a random variable following the Binomial law $\mathcal{B}(n,p)$ with $n = 100$ and $p = 0.01$. We obtain

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - 0.99^{100} \simeq 0.63.$$

There are other examples, where knowing the result given by the theory of probability may be more crucial. Consider for example the Russian roulette. What is the probability to be alive after three shots ? Consider also, – this time a realistic and psychologically painful situation –, the case of candidates to assisted procreation by in vitro fertilization. Considering that four attempts are reimbursed by the health insurance (in France), that the probability of success of each attempt is 20% and that (a disputable hypothesis) there is independence between each attempts, what is the probability to have a baby before being left on its own by the health insurance system? The answer is $\mathbb{P}(X \geq 1)$ where X follows a Binomial $\mathcal{B}(n,p)$ with $n = 4$ and $p = 0.2$. We compute

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - 0.8^4 \simeq 0.60,$$

which, one may think, is not that high.

2.6.2 Poisson's law.

Come back to the example of the production line in the previous paragraph about the Binomial law. In that example, n is large and p small, in such proportions that $np \sim 1$. This is also the case in the following cases:

– X= number of misprints in a book, the number of pages being $n = 300$ and the probability of a misprint in a page being $p = 0.01$ (assuming independence of misprints pages per pages),

– X=number of centenarian people in the French population divided by 10^4, the population being of $n.10^4$ people, the individual probability of being centenarian being $p = 3.10^{-4}$. 2016's data give $n = 6600$ then.

For such cases, is there a way to compute quite simply the probability $\mathbb{P}(X = k)$? The answer is given in Proposition 2.5.

Definition 2.19 (Poisson's law)**.** The Poisson's law is the law of a random variable X with values in $\mathbb{N}$, given by

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \tag{2.15}$$

where $\lambda > 0$ is a parameter. The Poisson's law of parameter λ is denoted by $\mathcal{P}(\lambda)$.

Proposition 2.5 (Convergence Binomial to Poisson's)**.** *For $\lambda > 0$, $n \in \mathbb{N}^*$ with $n > \lambda$, let X_n be a random variable of Binomial law $\mathcal{B}(n, \frac{\lambda}{n})$. Then we have the convergence*

$$\lim_{n\to+\infty} \mathbb{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \tag{2.16}$$

for each $k \in \mathbb{N}$.

Note that, in (2.16), we should write $\mathbb{P}_n(X_n = k)$, instead of $\mathbb{P}(X_n = k)$: the random variable X_n is defined on a probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and there is no reason to have the same probabilistic spaces for different indices. At the same time, it is also possible, by forming a countable product, to see all the random variables X_n defined on the same probability space. It is characteristic of a result of *convergence in law* that the probability space does not matter (see Section 2.8). The limit given in (2.16) is an instance of *convergence in law*: it says that the Binomial law $\mathcal{B}(n, \frac{\lambda}{n})$ can be approximated by $\mathcal{P}(\lambda)$ when $n \to +\infty$.

Proof of Proposition 2.5. We have

$$\begin{aligned}\mathbb{P}(X_n = k) &= \frac{n!}{(n-k)!\,k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}.\end{aligned}$$

The last term converges to $e^{-\lambda}\frac{\lambda^k}{k!}$ since two of the factors converges to 1 and $(1-\frac{\lambda}{n})^n$ converges to $e^{-\lambda}$. $\square$

Exercise 2.20 (Large Deviations)**.** This exercise is about the limit (in law) of the Binomial law in the regime "$n \to +\infty$, p fixed". If $X_n \sim \mathcal{B}(n,p)$, then we expect (see the Law of large numbers, Section 2.11) that $X_n \sim np$ for large n. Show that, for all $x \in (0,1)$,

$$\mathbb{P}(X_n = [xn]) = e^{-n[H(x;p)+o(1)]}, \tag{2.17}$$

when $[n \to +\infty]$, where $H(x;p)$ satisfies $H(x;p) > 0$ if $x \neq p$, $H(p;p) = 0$ (give the explicit expression of $H(x;p)$). In (2.17), $[y]$ is the integer part of y: the only integer $m \in \mathbb{N}$ such that $m \leq y < m+1$.

The solution to Exercise 2.20 is here.

2.6.3 Normal law.

For $X_n \sim \mathcal{B}(n,p)$, Exercise 2.20 shows that $\mathbb{P}(X_n = [xn])$ is exponentially small when $x \neq p$. This is a result on large deviations ("on the large deviations of X_n from its average pn", to state the sentence entirely). The following theorem is an instance of the Central Limit Theorem. It gives the asymptotic behaviour of the rescaled variable

$$Z_n = \frac{\sqrt{n}}{\sqrt{p(1-p)}}\left(\frac{X_n}{n} - p\right) = \frac{1}{\sqrt{np(1-p)}}(X_n - pn). \tag{2.18}$$

Definition 2.21 (Normal law)**.** Let $\sigma > 0$, $\mu \in \mathbb{R}$. A real-valued random variable X is said to follow the normal law $\mathcal{N}(\mu,\sigma^2)$ if

$$\mathbb{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y-\mu|^2}{2\sigma^2}} dy, \tag{2.19}$$

for all $a < b \in \mathbb{R}$.

Theorem 2.6 (Laplace – de Moivre's Theorem)**.** *Let $X_n \sim \mathcal{B}(n,p)$. Then the rescaled random variable Z_n defined by* (2.18) *converges in law to the normal law $\mathcal{N}(0,1)$:*

$$\mathbb{P}(a < Z_n < b) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy,$$

for all $a < b \in \mathbb{R}$, as $n \to +\infty$.

A proof of Theorem 2.6 is given in Section 2.10.2.

2.7 Expectancy

2.7.1 Integration of Banach-valued random variables

Let E be a separable Banach space endowed with the Borel σ-algebra $\mathcal{E}$. Let $X : \Omega \to E$ be a random variable. To define the integral

$$\mathbb{E}(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega),$$

we can apply Proposition 2.2: it gives the existence of a sequence (X_n) of simple functions which converges almost surely to X in E. Each integral $\mathbb{E}(X_n)$ is defined as the finite sum

$$\sum_{y\in X_n(\Omega)} \mathbb{P}(X_n^{-1}(\{y\}))\, y.$$

With this definition, and by the triangle inequality for finite sums, we have

$$\|\mathbb{E}(Y)\|_E \leq \mathbb{E}(\|Y\|_E) \tag{2.20}$$

for all Y simple. Assume

$$\mathbb{E}(\|X\|_E) < +\infty. \tag{2.21}$$

If (2.21) is realized, we say that X is *integrable*. When X is integrable, as $\|X_n - X\|_E \to 0$ almost surely and $\|X_n - X\|_E \leq 3\|X\|_E$ (this is due to the control of $\|X_n\|_E$ by $2\|X\|_E$ in Proposition 2.2), $\mathbb{E}(\|X_n - X\|_E)$ tends to 0 by the Dominated Convergence Theorem. Using (2.20) with $Y = X_n - X_m$, we deduce that

$$\|\mathbb{E}(X_n) - \mathbb{E}(X_m)\|_E \leq \mathbb{E}(\|X_n - X\|_E) + \mathbb{E}(\|X_m - X\|_E).$$

Therefore the sequence $(\mathbb{E}(X_n))$ is Cauchy in E and convergent to an element called $\mathbb{E}(X)$. If $(\widetilde{X}_n)$ is an other sequence of simple functions which converges almost surely to X in E and satisfies a uniform bound $\|\widetilde{X}_n\|_E \leq C\|X\|_E$, we obtain a second candidate $\widetilde{\mathbb{E}(X)}$ for the integral of X with respect to $\mathbb{P}$, but $\mathbb{E}(X) = \widetilde{\mathbb{E}(X)}$ since, by (2.20),

$$\|\mathbb{E}(X) - \widetilde{\mathbb{E}(X)}\|_E = \lim_{n\to+\infty} \|\mathbb{E}(X_n) - \mathbb{E}(\widetilde{X}_n)\|_E \leq \limsup_{n\to+\infty} \mathbb{E}(\|X_n - \widetilde{X}_n\|_E) = 0.$$

The integral $\mathbb{E}(X)$, called *expectancy* of X, is independent of the sequence of simple function which is used. Besides, the triangular inequality (2.20) is satisfied for $Y = X$.

Remark 2.6. For more details about the integration of Banach space valued functions (in the case where E is not separable in particular), see [Yos80, p. 130] or [Eva10, Appendix E].

Let us sum up some of the notions introduced here in the following definition.

Definition 2.22 (Integrable random variable)**.** Let E be a separable Banach space endowed with the Borel σ-algebra. Let $X\colon \Omega \to E$ be a random variable. We say that X is integrable if $\|X\|$ is. We denote by $L^1(\Omega; E)$ the set of integrable random variables $\Omega \to E$ modulo almost-sure equality.

Note also that we will keep the usual notation $L^1(\Omega)$ for $L^1(\Omega; \mathbb{R})$.

2.7.2 Expectancy, variance, independence

Definition 2.23 (Variance)**.** Let E be a separable Banach space endowed with the Borel σ-algebra $\mathcal{E}$. Let $X\colon\Omega\to E$ be a random variable such that

$$\mathbb{E}(\|X\|_E^2)<+\infty. \tag{2.22}$$

The *variance* of X is defined by $\mathbb{V}\mathrm{ar}(X)=\mathbb{E}\|X-\mathbb{E}(X)\|_E^2$.

If X has the Binomial distribution $\mathcal{B}(n,p)$, then

$$\mathbb{E}X=np,\quad \mathbb{V}\mathrm{ar}(X)=np(1-p). \tag{2.23}$$

The identities in (2.23) can be computed directly (X is a simple random variable, indeed), using (2.5), the formula

$$\mathbb{E}X=\sum_{k=0}^n k\mathbb{P}(X=k),\quad \mathbb{V}\mathrm{ar}(X)=\sum_{k=0}^n |k-np|^2\mathbb{P}(X=k)$$

and some variations (obtained by differentiation with respect to x) on the Binomial formula

$$\sum_{k=0}^n \binom{n}{k}x^k=(1+x)^n.$$

However, it is much easier to compute (2.23) by using the decomposition

$$X=X_1+\cdots+X_n,$$

where $X_1,\dots X_n$ are independent identically distributed Bernoulli random variables, with $\mathbb{P}(X_i=1)=p$, $\mathbb{P}(X_i=0)=1-p$. We have indeed $\mathbb{E}X_i=p$, $\mathbb{V}\mathrm{ar}(X_i)=p(1-p)$, hence

$$\mathbb{E}X=\sum_{i=1}^n\mathbb{E}X_i=np \tag{2.24}$$

and

$$\mathbb{V}\mathrm{ar}(X)=\sum_{i=1}^n\mathbb{V}\mathrm{ar}(X_i)=np(1-p). \tag{2.25}$$

The commutation between $\mathbb{E}$ and $\sum_{i=1}^n$ in (2.24) is a consequence of the linearity of $\mathbb{E}$. The commutation between $\mathbb{V}\mathrm{ar}$ and $\sum_{i=1}^n$ in (2.25) is *false* in general for the simple reason that $X\mapsto\mathbb{V}\mathrm{ar}(X)$ is quadratic. It is however *true* if we consider a sum of *independent* random variables. Indeed, we have, more generally, the following proposition.

Proposition 2.7 ($\mathbb{E}$ and independence)**.** *Let E_i, for $i=1,\dots,n$, be some separable Banach space endowed with the Borel σ-algebra $\mathcal{E}_i$. Let $X_i\colon\Omega\to E_i$ be some random variables and let $\phi_i\colon E_i\to\mathbb{R}$ be some measurable functions such that $\mathbb{E}|\phi_i(X_i)|<+\infty$ for every i. Assume that $(X_i)_{1,n}$ is* independent. *Then*

$$\mathbb{E}(\phi_1(X_1)\cdots\phi_n(X_n))=\mathbb{E}(\phi_1(X_1))\cdots\mathbb{E}(\phi_n(X_n)). \tag{2.26}$$

Proof of Proposition 2.7. We may reduce everything to the case $E_i = \mathbb{R}$, $\phi_i = \mathrm{Id}_{\mathbb{R}}$ be considering $Y_i = \phi_i(X_i)$. However, the identity (2.26) is more suggestive for the proof. Indeed, it is equivalent to

$$\int_{E_1\times\cdots\times E_n} f d\mu_{(X_1,\dots,X_n)} = \int_{E_1\times\cdots\times E_n} f d\mu_{X_1}\cdots d\mu_{X_n)},$$

for

$$f(x_1,\dots,x_n) = \phi_1(x_1)\cdots\phi_n(x_n).$$

Hence (2.26) follows from Theorem 2.3 and an argument of approximation for f. □

Recall in particular the formula

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \tag{2.27}$$

if X, Y are independent real-valued integrable random variable. The statement about the variance which we left aside is let as an exercise.

Exercise 2.24 (Linearity of the variance for independent random variables)**.** Let H be a separable Hilbert space endowed with the Borel σ-algebra. Let $X_1,\dots,X_n\colon\Omega\to H$ be some *independent* random variables satisfying the integrability condition (2.22) of order 2. Show that

$$\mathbb{V}\mathrm{ar}(X_1+\ldots+X_n) = \mathbb{V}\mathrm{ar}(X_1)+\ldots+\mathbb{V}\mathrm{ar}(X_n) \tag{2.28}$$

The solution to Exercise 2.24 is here.

Note that (2.27) is the identity that was lacking to complete the argument in the introductory paragraph Section 1. Indeed, taking expectation in (2.1), we obtain $\mathbb{E}|X_{N+1}|^2 = \mathbb{E}|X_N|^2 + 1$ since

$$\mathbb{E}\left[X_N Z_{N+1}\right] = \mathbb{E}[X_N]\mathbb{E}[Z_{N+1}] = 0 \tag{2.29}$$

by independence.

Note also the following fundamental identity, which occurred already in Proposition 2.7: if $\varphi\colon E\to\mathbb{R}$ is continuous and bounded (or more generally measurable and bounded), then

$$\mathbb{E}\varphi(X) = \int_E \varphi(x)d\mu_X(x) = \langle\mu_X,\varphi\rangle. \tag{2.30}$$

Indeed, (2.30) is true when $\varphi = \mathbf{1}_B$, B being a Borel subset of E. The general case follows by approximation.

2.8 Convergence in law

We have already encountered some examples of convergence in law (*cf.* Proposition 2.5 for example). We will be more specific about it in this paragraph. Our reference is *Convergence of probability measures*, by P. Billingsley, [Bil99].

Definition 2.25 (Weak convergence of probability measures)**.** Let E be a separable Banach space endowed with the Borel σ-algebra. Let μ_n, $n = 1, 2, \ldots$ be some Borel probability measures over E and let μ be a Borel measure on E. We say that (μ_n) converges weakly to μ (denoted $\mu_n \Rightarrow \mu$) if

$$\langle \varphi, \mu_n \rangle = \int_E \varphi d\mu_n \to \langle \varphi, \mu \rangle = \int_E \varphi d\mu, \tag{2.31}$$

for all continuous bounded function $\varphi \colon E \to \mathbb{R}$.

Note that the limit μ is then also a *probability* measure. This is a consequence of (2.31) with $\varphi \equiv 1$. Beware also that $\mu_n \Rightarrow \mu$ *does not* imply $\mu_n(A) \to \mu(A)$ for all Borel set A. The convergence $\mu_n(A) \to \mu(A)$ is true only if the limit measure μ does not charge the topological boundary ∂A, *i.e.* $\mu(\partial A) = 0$. In general, $\mu_n \Rightarrow \mu$ is equivalent to

$$\limsup_{n\to+\infty} \mu_n(F) \leq \mu(F), \text{ for all closed set } F, \tag{2.32}$$

and also equivalent to

$$\liminf_{n\to+\infty} \mu_n(G) \geq \mu(G), \text{ for all open set } G. \tag{2.33}$$

All the assertions above are part of the Portmanteau Theorem, [Bil99, Theorem 2.1]. There may be strict inequality in (2.32) and (2.33). This can be seen by considering $\mu_n = \delta_{x_n}$, where (x_n) is a sequence of points in E converging to an element $x \in \partial A$ ($A = F$ or G, depending on the characterization which is considered).

Definition 2.26 (Convergence in law)**.** Let E be a separable Banach space endowed with the Borel σ-algebra. Let

$$X_n \colon (\Omega_n, \mathcal{F}_n, \mathbb{P}_n) \to E, \quad X \colon (\Omega, \mathcal{F}, \mathbb{P}) \to E$$

be some random variables. We say that (X_n) *converges in law* to X (denoted $X_n \Rightarrow X$), if there is weak convergence of the laws: $\mu_{X_n} \Rightarrow \mu_X$. This means:

$$\mathbb{E}_n \varphi(X_n) \to \mathbb{E}\varphi(X), \tag{2.34}$$

for every continuous and bounded function $\varphi \colon E \to \mathbb{R}$.

Remark 2.7. The random variable X in Definition 2.26 is, in a way, superfluous. In essence, saying that (X_n) is converging in law means that there exists a Borel probability measure μ on E such that (μ_{X_n}) is converging weakly to μ. However, we can always find a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X such that $\mu = \mu_X$. We simply consider

$$(\Omega, \mathcal{F}, \mathbb{P}) = (E, \mathcal{B}(E), \mu), \quad X = \mathrm{Id}_E. \tag{2.35}$$

Nevertheless, we will often have the following situation: the random variables X_n are defined on the same probability space $(\Omega_*, \mathcal{F}_*, \mathbb{P}_*)$ and the sequence (μ_{X_n}) is converging weakly to a Borel probability measure μ on E. In that case, we would like to find a random variable X defined on $(\Omega_*, \mathcal{F}_*, \mathbb{P}_*)$ of law μ (this allows to write $\mathbb{P}_*(X_n \in A) \to \mathbb{P}_*(X \in A)$ for example, provided $\mathbb{P}_*(X \in \partial A) = 0$). Can we find such an X? It is not really necessary to answer to that question. To have a unique probability space, what we do instead is that we keep (2.35) and consider the the probability space

$$(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) = (\Omega_* \times \Omega, \mathcal{F}_* \times \mathcal{F}, \mathbb{P}_* \times \mathbb{P}).$$

Then we define $\tilde{X}_n(\omega_*, \omega) = X_n(\omega_*)$, $\tilde{X}(\omega_*, \omega) = X(\omega)$. Then $\mu_{\tilde{X}_n} = \mu_{X_n}$ and $\mu_{\tilde{X}} = \mu_X = \mu$.

Remark 2.8. Let us insist on the fact that only the laws of the random variables matter when considering convergence in law. For example, if X has the Bernoulli distribution $b(1/2)$, $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$ then $Y = 1 - X$ also. The sequence (X_n) defined by $X_{2n} = X$, $X_{2n+1} = Y$ is stationary (hence convergent) in law, but not convergent almost surely since it has two almost-sure convergent subsequences with distinct limits. We can modify the random variables, without affecting their distributions, to ensure convergence almost-sure, simply by setting $\tilde{X}_n = X$ for all n. This is an instance of the Skorohod representation theorem (Theorem 2.14 below) to which we will arrive ultimately in this Section 2.8. The following exercise provides an other instance of the Skorohod representation theorem.

Exercise 2.27. Let (X_n) be the sequence defined in Exercise 2.16.

1. Show that (X_n) is converging in law to a limit X.
2. Build a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and some random variables $\tilde{X}_n$, $\tilde{X}$ on $\tilde{\Omega}$ such that
   - for all $n \in \mathbb{N}^*$, the random variables $\tilde{X}_n$ and X_n have the same law; $\tilde{X}$ and X have the same law,
   - $(\tilde{X}_n)$ is converging to $\tilde{X}$ $\tilde{\mathbb{P}}$-almost-surely.

The solution to Exercise 2.27 is here.

Proposition 2.8. *Let E be a separable Banach space endowed with the Borel σ-algebra. Let μ_n, $n = 1, 2, \dots$ be some Borel probability measures over E and let μ be a Borel probability measure on E. For the weak convergence of (μ_n) to μ it is sufficient that $\langle \mu_n, \varphi \rangle \to \langle \mu, \varphi \rangle$ for all* uniformly continuous *and bounded function φ on E.*

Proof of Proposition 2.8. We use the criterion (2.32). Let $F \subset E$ be a closed set. Let d be the metric[1] induced by the norm on E and let $d(x, F) = \inf_{y \in E} d(x, y)$ denote the distance to E. The sequence of functions

$$\varphi_k \colon x \mapsto (1 - k d(x, F))^+ \tag{2.36}$$

[1]actually, in all this paragraph, we may have assumed E to be a metric space, see [Bil99]

is increasing and tends to $\mathbf{1}_F$. The function φ_k is also uniformly continuous (it is even k-Lipschitz continuous). If $\langle \mu_n, \varphi_k\rangle \to \langle \mu, \varphi_k\rangle$ for each k, we have therefore

$$\mu_n(F) \leq \langle \mu_n, \varphi_k\rangle \to \langle \mu, \varphi_k\rangle$$

and thus $\limsup_{n\to+\infty} \mu_n(F) \leq \langle \mu, \varphi_k\rangle$. At the limit $k \to +\infty$, we obtain (2.32). □

Remark 2.9. The function φ_k in (2.36) can also be defined as

$$\varphi_k(x) = \sup_{y\in E} \left[\varphi(y) - k\|x-y\|_E\right], \quad \varphi(x) = \mathbf{1}_F(x). \tag{2.37}$$

Formula (2.37) defines the sup-convolution of the function φ.

In the proof of Proposition 2.8, two elements appear, in a more or less transparent way:

- the regularisation of functions in infinite dimension, as already emphasized in Remark 2.9,
- the fact that $C_b(E)$ is a separating class.

These two elements are explained in more details in Proposition 2.9 and Proposition 2.10 below.

Proposition 2.9. *Let E be a separable Banach space. Let φ be a continuous, bounded function on E. There is a sequence (φ_n) of Lipschitz continuous bounded functions on E such that*

$$\sup_n \sup_{x\in E} |\varphi_n(x)| \leq \sup_{x\in E} |\varphi(x)|, \quad \varphi_n(x) \to \varphi(x), \tag{2.38}$$

for all $x \in E$.

Proof of Proposition 2.9. Let $\varphi \in C_b(E)$ (continuous, bounded function). Without loss of generality, we assume $0 \leq \varphi \leq 1$. For $n \in \mathbb{N}$, we consider the inf-convolution φ_n of φ defined by

$$\varphi_n(x) = \inf_{y\in E} \left[\varphi(y) + n\|x-y\|_E\right]. \tag{2.39}$$

Taking $y = x$ in (2.39), we see that $0 \leq \varphi_n(x) \leq \varphi(x) \leq 1$. From the triangular inequality, one deduces that φ_n is n-Lipschitz continuous. Let $\varepsilon > 0$, and let $y_{n,\varepsilon} \in E$ be such that

$$\varphi_n(x) - \varepsilon \leq \varphi(y_{n,\varepsilon}) + n\|x - y_{n,\varepsilon}\|_E \leq \varphi_n(x) + \varepsilon.$$

We have then

$$n\|x - y_{n,\varepsilon}\|_E \leq \varphi(y_{n,\varepsilon}) + n\|x - y_{n,\varepsilon}\|_E \leq \varphi_n(x) + \varepsilon \leq 1 + \varepsilon.$$

Therefore $y_{n,\varepsilon} \to x$ when $n \to +\infty$. For n large enough we have therefore, by continuity of φ (lower semi-continuity is sufficient actually),

$$\varphi(x) \leq \varphi(y_{n,\varepsilon}) + \varepsilon \leq \varphi(y_{n,\varepsilon}) + n\|x - y_{n,\varepsilon}\|_E + \varepsilon \leq \varphi_n(x) + 2\varepsilon.$$

Eventually, we obtain $\varphi(x) - 2\varepsilon \leq \varphi_n(x) \leq \varphi(x)$. This shows that $\varphi_n(x) \to \varphi(x)$. □

We denote by $C_b(E)$ the set of continuous, bounded functions on E and by $\mathrm{Lip}\cap C_b(E)$ the subset of Lipschitz continuous, bounded functions.

Definition 2.28. Let E be a separable Banach space. A subset X of the set of bounded measurable functions $E \to \mathbb{R}$ is said to be a *separating class* if two Borel probability measures that coincide on X are equal.

Proposition 2.10. *Let E be a separable Banach space. Then* $\mathrm{Lip}\cap C_b(E)$ *is a separating class.*

Proof of Proposition 2.10. Let μ and ν be some Borel probability measures on E such that $\langle\varphi,\mu\rangle = \langle\varphi,\nu\rangle$ for all $\varphi \in \mathrm{Lip}\cap C_b(E)$. We want to show that $\mu(A) = \nu(A)$ for all Borel subset A of E. The measures μ and ν are inner regular, [Bil99, Theorem 1.1]: $\mu(A) = \sup \mu(F)$, where the sup is taken over closed subsets F of A. Consequently, it is sufficient to consider the case A closed. We have shown (Remark 2.9) that $\mathbf{1}_A$ is the simple limit of a sequence of Lipschitz bounded functions φ_n. By dominated convergence (or monotone convergence, since $n \mapsto \varphi_n$ is monotone), we have

$$\mu(A) = \langle \mathbf{1}_A, \mu\rangle = \lim_{n\to+\infty} \langle\varphi_n,\mu\rangle = \lim_{n\to+\infty}\langle\varphi_n,\nu\rangle = \nu(A).$$

This gives the result. □

2.8.1 Convergence in probability

Definition 2.29 (Convergence in probability)**.** Let E be a separable Banach space endowed with the Borel σ-algebra. A sequence (X_n) of random variables on E is said to converge *in probability* to a random variable X if, for all $\delta > 0$,

$$\mathbb{P}(\|X_n - X\|_E > \delta) \to 0, \tag{2.40}$$

when $n \to +\infty$.

Note that (2.40) can also be written

$$\mathbb{E}\mathbf{1}_{\|X_n - X\|_E > \delta} \to 0, \tag{2.41}$$

when $n \to +\infty$. Since $\mathbf{1}_{\|X_n-X\|_E>\delta}$ is bounded by the constant, integrable function $\mathbf{1}$, almost-sure convergence implies convergence in probability by the Dominated convergence theorem. Convergence in probability implies convergence in law.

Proposition 2.11. *Let E be a separable Banach space endowed with the Borel σ-algebra. Let (X_n) be a sequence of random variables on E which converges in probability to a random variable X. Then (X_n) is converging in law to X.*

Proof of Proposition 2.11. By Proposition 2.8, it is sufficient to show that $\mathbb{E}\varphi(X_n) \to \mathbb{E}\varphi(X)$ for $\varphi \in C_b(E)$ uniformly continuous. If φ is *uniformly* continuous, with a modulus of continuity denoted by ω_φ, then the conclusion comes from the following estimate: we bound the difference $|\mathbb{E}\varphi(X_n) - \mathbb{E}\varphi(X)|$ by the sum of the two terms

$$\begin{aligned}&\mathbb{E}\left[|\varphi(X_n) - \varphi(X)|\mathbf{1}_{\|X_n - X\|_E > \delta}\right] + \mathbb{E}\left[|\varphi(X_n) - \varphi(X)|\mathbf{1}_{\|X_n - X\|_E \leq \delta}\right] \\ &\quad \leq \|\varphi\|_{C_b(E)}\mathbb{P}(\|X_n - X\|_E > \delta) + \omega_\varphi(\delta),\end{aligned}$$

where $\|\varphi\|_{C_b(E)} = \sup_{x \in E}|\varphi(x)|$. The right-hand side can be made arbitrary small by choosing first δ small, then n large. □

Lemma 2.12. *Let E be a separable Banach space endowed with the Borel σ-algebra. Let (X_n), (Y_n) be some sequences of random variables on E such that (X_n) converges in law to a random variable X and $X_n - Y_n$ converges to 0 in probability. Then (Y_n) is converging in law to X.*

Exercise 2.30. Give the proof of Lemma 2.12.
The solution to Exercise 2.30 is here.

Proof of Lemma 2.12. we use the characterization (2.32) of convergence in law. Let F be a closed subset of E. Let $\varepsilon > 0$ and $\delta > 0$. There exists an n_0 such that $\mathbb{P}(\|X_n - Y_n\|_E > \delta) < \varepsilon$ for all $n \geq n_0$. We have then $\mathbb{P}(Y_n \in F) < \varepsilon + \mathbb{P}(X_n \in \overline{F}^\delta)$, where $\overline{F}^\delta$ denotes the δ-neighbourhood of F:

$$\overline{F}^\delta = \{x \in E; d(x, F) \leq \delta\}, \quad d(x, F) = \min_{y \in F}\|x - y\|_E.$$

Since $\overline{F}^\delta$ is closed, we obtain

$$\limsup_{n \to +\infty} \mathbb{P}(Y_n \in F) \leq \varepsilon + \mu_X(\overline{F}^\delta).$$

Since $(\overline{F}^\delta) \downarrow F$ when $\delta \downarrow 0$ (because F is closed), we obtain $\limsup_{n \to +\infty} \mathbb{P}(Y_n \in F) \leq \varepsilon + \mu_X(F)$ at the limit $\delta \to 0$. Since ε is arbitrary, this gives the result. □

Exercise 2.31. Let (X_n) be a sequence of random variables and X, Y random variables such that (X_n, X) converges in law to (Y, Y) (*i.e.* $\mu_{(X_n, X)}$ is converging weakly to a probability measure concentrated on the diagonal of $E \times E$). Show that (X_n) is converging in probability to X.
The solution to Exercise 2.31 is here.

2.8.2 Prohorov's theorem and Skorohod's representation theorem

Definition 2.32 (Tightness)**.** Let E be a separable Banach space endowed with the Borel σ-algebra. A family $\mathfrak{P}$ of Borel probability measures over E is said to be *tight* if, for every $\varepsilon > 0$, there exists a compact $K \subset E$ such that

$$\mu(K) \geq 1 - \varepsilon,$$

for all $\mu \in \mathfrak{P}$.

Exercise 2.33. Show that the following families are tight.

1. $\mathfrak{P} = \{\mu\}$ (a single element), $E = \mathbb{R}$,

2. $\mathfrak{P} = \{\mu\}$, E σ-compact,

3. $\mathfrak{P} = \{\mu\}$, E a separable Banach space,

4. $E = L^2(\mathbb{T}^d)$ ($\mathbb{T}^d$ is the d-dimensional torus), $\mathfrak{P} = \{\mu_n; n \in \mathbb{N}\}$, where μ_n is the law of a random variable X_n satisfying the estimate

$$\sup_{n\in\mathbb{N}} \mathbb{E}\|X_n\|_{H^1(\mathbb{T}^d)} < +\infty, \tag{2.42}$$

   where $H^1(\mathbb{T}^d)$ is the Sobolev space H^1 over $\mathbb{T}^d$.

5. Generalize the preceding example to separable Banach spaces E, F with compact injection of F into E.

6. *(Reflected random walk)* Consider the reflected random walk defined on Figure 1. Let X_n be the position at time n. We assume $X_0 = 0$ (the choice of X_0 is not

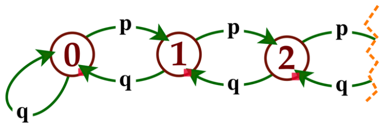


Figure 1: Reflected random walk; $p + q = 1$

   relevant here). Show that the family $\{\mu_{X_n}; n \in \mathbb{N}\}$ is tight if, and only if, $p < \frac{1}{2}$.

The solution to Exercise 2.33 is here.

Theorem 2.13 (Prohorov's theorem)**.** *Let E be a separable Banach space endowed with the Borel σ-algebra. Let μ_n, $n = 1, 2, \dots$ be some Borel probability measures over E. Then there is equivalence between:*

1. *each subsequence of (μ_n) admits a subsequence converging weakly,*

2. *the family $\{\mu_n; n \in \mathbb{N}\}$ is tight.*

There are different ways to put a metric on the set $\mathcal{P}_1(E)$ of Borel probability measures on E, which turn $\mathcal{P}_1(E)$ into a separable, complete metric space in which convergent sequences are sequences converging weakly (see, for instance [Bil99, page 72] on the Prohorov metric). In that context, the Prohorov theorem may be rephrased as follows: a set $\mathfrak{P}$ of Borel probability measures on E is relatively compact if, and only if, it is tight.

Theorem 2.14 (Skorohod's representation theorem)**.** *Let E be a separable Banach space endowed with the Borel σ-algebra. Let (X_n) be a sequence of random variables which converges in law to a random variable X. Then there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and some random variables $\tilde{X}_n$, $\tilde{X}$ on $\tilde{\Omega}$ such that*

1. *for all $n \in \mathbb{N}^*$, the random variables $\tilde{X}_n$ and X_n have the same law; $\tilde{X}$ and X have the same law,*
2. *$(\tilde{X}_n)$ is converging to $\tilde{X}$, $\tilde{\mathbb{P}}$-almost-surely.*

See [Bil99, p. 60] for the proof of the Prohorov theorem and [Bil99, p. 70] for the proof of the Skorohod theorem.

2.9 Conditional expectancy

Theorem-Definition 2.15 (Conditional expectancy)**.** *Let E be a separable Banach space endowed with the Borel σ-algebra $\mathcal{E}$. Let $X\colon (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ be a random variable. Assume X to be integrable and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then there exists a unique $\mathcal{G}$-measurable random variable in $L^1(\Omega, \mathcal{G}; E, \mathcal{E})$, denoted $\mathbb{E}(X|\mathcal{G})$ and called the* conditional expectancy *of X knowing $\mathcal{G}$, such that*

$$\mathbb{E}\left[\mathbf{1}_A X\right] = \mathbb{E}\left[\mathbf{1}_A \mathbb{E}(X|\mathcal{G})\right], \tag{2.43}$$

for all $A \in \mathcal{G}$.

The random variable $\mathbb{E}(X|\mathcal{G})$ should be understood as the average of X with respect to "all that is not $\mathcal{G}$". This principle is illustrated by the following examples.

Example 2.10. Take $\mathcal{G} = \{\emptyset, \Omega\}$. Then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ a.s.

Example 2.11. Take $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$ where $B \in \mathcal{F}$. Then

$$\mathbb{E}(X|\mathcal{G}) = \frac{\mathbb{E}(\mathbf{1}_B X)}{\mathbb{P}(B)} \mathbf{1}_B + \frac{\mathbb{E}(\mathbf{1}_{B^c} X)}{\mathbb{P}(B^c)} \mathbf{1}_{B^c} \text{ a.s.}$$

Example 2.12. If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then

$$\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}) \text{ a.s.} \tag{2.44}$$

Example 2.13. One has

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X). \tag{2.45}$$

Example 2.14. Let X and Y be some *independent* random variables, $\Phi\colon E \times E \to \mathbb{R}$ continuous and bounded. Then

$$\mathbb{E}(\Phi(X,Y)|\sigma(X)) = f(X) \text{ a.s.}, \quad f(x) := \mathbb{E}\Phi(x, Y). \tag{2.46}$$

Exercise 2.34 (Examples of conditional expectancies)**.** Prove the assertions in Example 2.10 to 2.14.

The solution to Exercise 2.34 is here.

The identities (2.45) and (2.46) are fundamental (see the treatment of *Example 1* page 52 in particular). For example, let us prove that the random walk (X_n) defined in Section 1 is a *Markov chain*. This means, roughly speaking that, conditionally to the knowledge of the past up to time n, it is actually sufficient to know the present state at time n to determine the state at time $n+1$. The exact mathematical condition is the following one.

Definition 2.35 (Markov chain)**.** Let E be a separable Banach space endowed with the Borel σ-algebra. A sequence of random variables $X_0, X_1, \dots$ over E is said to be a *Markov chain* if, for all $n \geq 0$,

$$\mathbb{E}(\phi(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(\phi(X_{n+1})|\sigma(X_n)) \text{ a.s.}, \tag{2.47}$$

for all continuous and bounded $\phi \colon E \to \mathbb{R}$, where $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ is the σ-algebra of "the past up to time n" and $\sigma(X_n)$ is the σ-algebra of "the present state at time n".

To prove (2.47) for the random walk (X_n), apply (2.46) first, to obtain

$$\mathbb{E}(\phi(X_{n+1})|\sigma(X_n)) = \mathbb{E}(\phi(X_n + Z_{n+1})|\sigma(X_n)) = \frac{1}{2}\left(\phi(X_n - 1) + \phi(X_n + 1)\right) \text{ a.s.}$$

Apply then (2.46) again to $\Phi \colon (z, y) \mapsto \phi(z \cdot \mathbf{1} + y)$, where $z = (z_i)_{1,n}$, $\mathbf{1}$ is the vector in $\mathbb{R}^n$ with all components 1. We take also $Y = Z_{n+1}$ and note that $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$ to get

$$\mathbb{E}(\phi(X_{n+1})|\mathcal{F}_n) = \frac{1}{2}\left(\phi(X_n - 1) + \phi(X_n + 1)\right) \text{ a.s.}$$

Exercise 2.36 (Example of Markov chain)**.** Let E, F be some separable Banach spaces, let f be a measurable bounded application $E \times F \to E$. Let (Y_n) be independent random variables in F and (X_n) the sequence defined by $X_0 = x \in E$, $X_{n+1} = f(X_n, Y_n)$. Show that (X_n) is a Markov chain, *i.e.* satisfies (2.47).

The solution to Exercise 2.36 is here.

The property (2.46) is also the central argument to prove that the transition operator (homogeneous case) is a semi-group (see the treatment of *Example 1* page 52 again, for example).

2.10 Quantitative convergence in law: Stein's method

We have seen in Proposition 2.5 an example of convergence in law "Binomial⇒Poisson". Theorem 2.6 is an instance of the Central Limit Theorem and gives a result of convergence in law "Binomial⇒Normal".

Exercise 2.37 (Laplace - de Moivre's Theorem)**.** Justify that Theorem 2.6 is indeed a result of convergence in law (you may use the characterization (2.33) for example).

The solution to Exercise 2.37 is here.

We will now give quantitative versions of those results. The general idea of Stein's method [Ste72], that we put quite informally here, is the following one. Let μ and ν be some Borel probability measures on a separable Banach space E and let X and Y be some random variables of law μ and ν respectively. Let L be an operator acting on functions such that

$$\langle \nu, L\varphi \rangle = 0 \text{ for all } \varphi \iff \nu = \mu. \tag{2.48}$$

In (2.48), φ is a function $E \to \mathbb{R}$ with a certain regularity which we do not specify at the moment. If we define the operator L^* by duality, $\langle L^*\nu, \varphi \rangle := \langle \nu, L\varphi \rangle$, then (2.48) can be written more concisely as $\mathrm{Ker}(L^*) = \{\mu\}$. For a given function ψ, we expect[2] then the equation

$$L\varphi = \psi - \langle \mu, \psi \rangle \tag{2.49}$$

to be solvable, and the solution φ to be estimated by ψ (one has to specify the norms at that point). We apply ν to both members of (2.49) then and obtain

$$\langle \nu, \psi \rangle - \langle \mu, \psi \rangle = \langle \nu, L\varphi \rangle. \tag{2.50}$$

An estimate on $\langle \nu, L\varphi \rangle$ will indicate how close $\langle \nu, \psi \rangle$ is from $\langle \mu, \psi \rangle$.

The method was originally developed by Stein for the approximation of Gaussian random variable, [Ste72]. It was adapted by Chen in 1975, [Che75], to estimate the approximation of Poisson's distribution. This is the example we will treat first.

2.10.1 Convergence to the Poisson distribution

Let $Y_1, Y_2, \ldots$ be some independent random variables of Bernoulli's law $b(p_1), b(p_2), \ldots$: $\mathbb{P}(Y_i = 1) = p_i = 1 - \mathbb{P}(Y_i = 0)$. Let

$$X_n = Y_1 + \cdots + Y_n$$

and let X be a random variable of Poisson's distribution of parameter λ. We have seen in Proposition 2.5 that $X_n \Rightarrow X$ if $p_i = \frac{\lambda}{n}$. More generally, we have the following result.

Theorem 2.16 (Convergence Binomial to Poisson's)**.** *Let $Y_1, Y_2, \ldots$ be some independent random variables of Bernoulli's law $b(p_1), b(p_2), \ldots$: $\mathbb{P}(Y_i = 1) = p_i = 1 - \mathbb{P}(Y_i = 0)$. Let*

$$X_n = Y_1 + \cdots + Y_n,$$

and let X be a random variable of Poisson's distribution of parameter

$$\lambda = \sum_{i=1}^{n} p_i. \tag{2.51}$$

[2] By analogy with the Fredholm's alternative, *cf.* [Eva10, Th. 4 p. 321] in the context of elliptic equations, which may be summed up as $\mathrm{Im}(L) \simeq \mathrm{Ker}(L^*)^{\perp}$

We have the estimate in total variation distance

$$d_{\mathrm{TV}}(\mu_{X_n}, \mu_X) \le C\left(1 \wedge \frac{1}{\lambda}\right) \sum_{i=1}^{n} p_i^2, \tag{2.52}$$

where the total variation distance between two probability measures μ and ν on $\mathbb{R}$ is defined as the supremum of $|\mu(A) - \nu(A)|$ over Borel sets A in $\mathbb{R}$.

In the case $p_i = \frac{\lambda}{n}$, X_n has the distribution $\mathcal{B}(n, \lambda n^{-1})$ and (2.52) reads

$$d_{\mathrm{TV}}(\mu_{X_n}, \mu_X) \le C\lambda\,(1 \wedge \lambda)\,\frac{1}{n}.$$

We obtain the convergence $\mu_{X_n} \Rightarrow \mu_X$ stated in Proposition 2.5, together with an estimate of the distance of μ_{X_n} to μ_X.

Proof of Theorem 2.16. A probability measure ν over $\mathbb{N}$ has the decomposition

$$\nu = \sum_{x\in\mathbb{N}} \nu_x \delta_x, \quad \nu_x := \nu(\{x\}).$$

What characterizes the Poisson's law $\mu = \mathcal{P}(\lambda)$ is the relation $\mu_x = \frac{\lambda}{x}\mu_{x-1}$ for $x \ge 1$ (other recurrence relations, possibly more elaborate ones, are possible of course). Using the constraint

$$\sum_{x\in\mathbb{N}} \mu_x = 1,$$

we have $\mathrm{Ker}(L^*) = \{\mu\}$, where $(L^*\nu)_x = \lambda\nu_{x-1} - x\nu_x$ for $x \ge 1$, $(L^*\nu)_0 = 0$. We compute then

$$L\psi(x) = \lambda\psi(x+1) - x\psi(x). \tag{2.53}$$

For $\psi\colon \mathbb{N} \to \mathbb{R}$ bounded, set $\bar\psi(x) = \psi(x) - \langle \mu_X, \psi\rangle$ and let φ solve (2.49). One can compute

$$\varphi(x+1) = \lambda^{-x-1} x! \sum_{y=0}^{x} \frac{\lambda^y}{y!}\bar\psi(y) = -\lambda^{-x-1} x! \sum_{y=x+1}^{\infty} \frac{\lambda^y}{y!}\bar\psi(y),$$

for $x \ge 0$ (note that the value $\varphi(0)$ is irrelevant here). We will admit the following (non-trivial) result [BE83]:

$$\|\varphi\|_\infty \le \left[1 \wedge \frac{1.4}{\lambda}\right] \|\psi\|_\infty, \quad \|\Delta\varphi\|_\infty \le \left[1 \wedge \frac{1}{\lambda}\right] \|\psi\|_\infty, \tag{2.54}$$

where $\|\varphi\|_\infty = \sup_{x\in\mathbb{N}} |\varphi(x)|$, $\Delta\varphi(x) = \varphi(x+1) - \varphi(x)$.

We use (2.54) as follows: let ψ be the characteristic function of a set $A \in \mathbb{N}$. By (2.50), we have

$$\mu_{X_n}(A) - \mu_X(A) = \langle \mu_{X_n}, L\varphi\rangle = \mathbb{E}\left[\lambda\varphi(X_n+1) - X_n\varphi(X_n)\right].$$

Using the definition (2.51) of λ gives us

$$\mu_{X_n}(A) - \mu_X(A) = \sum_{i=1}^{n} \mathbb{E}\left[p_i\varphi(X_n+1) - Y_i\varphi(X_n)\right],$$

since $X_n = Y_1 + \cdots + Y_n$. Set

$$X_n^{(i)} = Y_1 + \cdot + \hat{Y}_i + \cdots + Y_n = X_n - Y_i.$$

Note that $X_n^{(i)}$ is independent on Y_i. Conditioning on Y_i, we have therefore

$$\mu_{X_n}(A) - \mu_X(A) = \sum_{i=1}^{n} \mathbb{E}\left[p_i(1-p_i)\varphi(X_n^{(i)}+1) + p_i(p_i\varphi(X_n^{(i)}+2) - \varphi(X_n^i+1))\right] \tag{2.55}$$

$$= \sum_{i=1}^{n} p_i^2 \mathbb{E}\Delta\varphi(X_n^{(i)}+1).$$

By (2.54), we obtain (2.52). □

Exercise 2.38 (End of the proof)**.** Justify (2.55).

The solution to Exercise 2.38 is here.

2.10.2 Quantitative CLT

Definition 2.39 (Monge-Kantorovitch distance W_1)**.** Let $\mathcal{P}_1(\mathbb{R})$ be the set of Borel probability measures ν on $\mathbb{R}$ having finite first moment:

$$\int_{\mathbb{R}} |x| d\nu(x) < +\infty.$$

The Monge-Kantorovitch distance $W_1(\mu,\nu)$ of two probability measures $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ is defined as the supremum of

$$\langle \mu - \nu, \psi \rangle = \int_{\mathbb{R}} \psi d\mu - \int_{\mathbb{R}} \psi d\nu$$

over all 1-Lipschitz continuous functions $\psi\colon \mathbb{R} \to \mathbb{R}$.

Any $\psi\colon \mathbb{R} \to \mathbb{R}$ which is 1-Lipschitz continuous is also sub-linear: $|\psi(x)| \le |x| + |\psi(0)|$. This shows that $\langle \psi, \nu \rangle$ is well defined for $\nu \in \mathcal{P}_1(\mathbb{R})$. Actually, rewriting

$$\langle \mu - \nu, \psi \rangle = \int_{\mathbb{R}} \int_{\mathbb{R}} (\psi(x) - \psi(y)) d\mu(x) d\nu(y),$$

we obtain the following bound

$$|\langle \mu - \nu, \psi \rangle| \le \int_{\mathbb{R}} |x| d\mu(x) + \int_{\mathbb{R}} |x| d\nu(x),$$

which is independent on ψ.

Theorem 2.17 (Quantitative CLT)**.** *Let $X_1, X_2, \dots$ be some independent, identically distributed random variables on $\mathbb{R}$ satisfying*

$$\mathbb{E}|X_n|^3 < +\infty, \quad \mathbb{E}X_n = 0, \quad \mathbb{V}\mathrm{ar}(X_n) = 1. \tag{2.56}$$

Define the renormalized sum

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}. \tag{2.57}$$

Then

$$W_1(\mu_{Z_n}, \mu) \leq \frac{3\mathbb{E}|X_1|^3}{\sqrt{n}}, \tag{2.58}$$

where μ is the normal law $\mathcal{N}(0,1)$.

In (2.56), only the hypothesis on the third moment matters. If the X_n's are not centred and reduced, say

$$\mathbb{E}X_n = \mu, \quad \mathbb{V}\mathrm{ar}(X_n) = \sigma^2 > 0,$$

then the result applies with the renormalized random variable

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}. \tag{2.59}$$

Indeed, the rescaling of the sum $X_1 + \cdots + X_n$ in Z_n is such that $\mathbb{E}Z_n = 0$ (by linearity of the operator $\mathbb{E}$) and $\mathbb{V}\mathrm{ar}(Z_n) = 1$ (by linearity of $\mathbb{V}\mathrm{ar}$ on sums of independent random variables and the property $\mathbb{V}\mathrm{ar}(\alpha X) = \alpha^2 \mathbb{V}\mathrm{ar}(X)$).

Proof of Theorem 2.17. Let μ denote the normal law $\mathcal{N}(0,1)$: μ has the density

$$\gamma(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

with respect to the Lebesgue measure on $\mathbb{R}$. Let $L\varphi$ be defined by

$$L\varphi(z) = \varphi'(z) - z\varphi(z). \tag{2.60}$$

Integration by parts shows that $\langle \mu, L\varphi \rangle = 0$ for all $\varphi \in C_b^1(\mathbb{R})$ (C^1 functions $\mathbb{R} \to \mathbb{R}$, bounded with bounded first derivative).

Let $\psi \in C_b^1(\mathbb{R})$ (C^1 functions $\mathbb{R} \to \mathbb{R}$, bounded with bounded first derivative), and $\bar{\psi}(z) := \psi(z) - \langle \mu, \psi \rangle$ then

$$\varphi(z) = e^{z^2/2} \int_{-\infty}^{z} e^{-r^2/2} \bar{\psi}(r) dr \tag{2.61}$$

$$= -e^{z^2/2} \int_{z}^{+\infty} e^{-r^2/2} \bar{\psi}(r) dr \tag{2.62}$$

(the identity (2.61)=(2.62) is due to $\langle \mu, \bar{\psi} \rangle = 0$) is in $C^2(\mathbb{R})$ and satisfies the equation $L\varphi = \bar{\psi}$. Admit for a moment the following result

Lemma 2.18. *The function φ defined by (2.61) is in $C_b^2(\mathbb{R})$ and satisfies the bounds*

$$\|\varphi\|_\infty \leq \|\psi'\|_\infty, \quad \|\varphi'\|_\infty \leq 4\|\psi\|_\infty, \quad \|\varphi''\|_\infty \leq 4\|\psi'\|_\infty, \tag{2.63}$$

where $\|\cdot\|_\infty$ is the sup-norm: $\|\varphi\|_\infty = \sup_{z\in\mathbb{R}} |\varphi(z)|$.

In virtue of Lemma 2.18, a measure $\nu \in \mathcal{P}_1(\mathbb{R})$ satisfying $\langle \nu, L\varphi\rangle = 0$ for all $\varphi \in C_b^1(\mathbb{R})$ is equal to μ, *i.e.* $\mathrm{Ker}(L^*) = \{\mu\}$. Indeed, taking φ solution to $L\varphi = \bar{\psi}$, we obtain

$$0 = \langle \nu, L\varphi\rangle = \langle \nu, \psi\rangle - \langle \mu, \psi\rangle\langle \nu, \mathbf{1}\rangle = \langle \nu, \psi\rangle - \langle \mu, \psi\rangle,$$

for all $\psi \in C_b^1(\mathbb{R})$, and this yields $\nu = \mu$. Therefore L is a good operator for Stein's method. Let us now complete the proof of Theorem 2.17. Let $\bar{X}_i = \frac{1}{\sqrt{n}} X_i$. Let

$$Z_n^{(i)} = \bar{X}_1 + \cdots + \widehat{\bar{X}_i} + \cdots + \bar{X}_n = Z_n - \bar{X}_i.$$

We decompose

$$\mathbb{E}L\varphi(Z_n) = \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\bar{X}_i{}^2 \varphi'(Z_n^{(i)} + \bar{X}_i) - \bar{X}_i\varphi(Z_n^{(i)} + \bar{X}_i)\right].$$

Conditioning on $\sigma(\bar{X}_i)$, we obtain (*cf.* (2.46)) $\mathbb{E}L\varphi(Z_n) = \sum_{i=1}^n \mathbb{E}\varphi_i(\bar{X}_i)$, where

$$\varphi_i(x) = \mathbb{E}\bar{X}_i{}^2 \mathbb{E}\varphi'(Z_n^{(i)} + x) - x\mathbb{E}\varphi(Z_n^{(i)} + x).$$

We use the Taylor formula

$$\varphi(Z_n^{(i)} + x) = \varphi(Z_n^{(i)}) + x\varphi'(Z_n^{(i)}) + \frac{x^2}{2}\varphi''(\cdot), \quad \varphi'(Z_n^{(i)} + x) = \varphi'(Z_n^{(i)}) + x\varphi''(\cdot),$$

to obtain

$$\begin{aligned}\varphi_i(\bar{X}_i) =& (\mathbb{E}\bar{X}_i{}^2 - \bar{X}_i{}^2)\mathbb{E}\varphi'(Z_n^{(i)}) - \bar{X}_i\mathbb{E}\varphi(Z_n^{(i)}) \\ &+ \bar{X}_i\mathbb{E}\bar{X}_i{}^2\mathbb{E}\varphi''(\cdot) - \frac{\bar{X}_i{}^3}{2}\mathbb{E}\varphi''(\cdot).\end{aligned} \tag{2.64}$$

Taking expectation, the first line (2.64) vanishes. Since $\mathbb{E}|\bar{X}_i|\mathbb{E}\bar{X}_i{}^2 \leq \mathbb{E}|\bar{X}_i|^3$ by the Hölder inequality, we obtain

$$|\mathbb{E}\varphi_i(\bar{X}_i)| \leq \frac{3}{2}\mathbb{E}|\bar{X}_i|^3\|\varphi''\|_\infty$$

and thus

$$|\mathbb{E}L\varphi(Z_n)| \leq \frac{3n}{2}\mathbb{E}|\bar{X}_i|^3\|\varphi''\|_\infty = \frac{3}{2\sqrt{n}}\mathbb{E}|X_i|^3\|\varphi''\|_\infty.$$

We use the last inequality in (2.63) to conclude. □

Proof of Lemma 2.18. For $z \leq 0$, (2.61) gives

$$|z\varphi(z)| \leq e^{z^2/2} \int_{-\infty}^{z} |r| e^{-r^2/2} dr \|\bar{\psi}\|_\infty = \|\bar{\psi}\|_\infty \leq 2\|\psi\|_\infty. \tag{2.65}$$

We obtain the same bound for $z \geq 0$ thanks to (2.62). Using the equation $L\varphi = \bar{\psi}$, we deduce $\|\varphi'\|_\infty \leq 4\|\psi\|_\infty$. Note that the estimate can be improved and that

$$\|\varphi'\|_\infty \leq 2\|\psi\|_\infty \text{ if } \langle \psi, \mu \rangle = 0. \tag{2.66}$$

Let us now prove that

$$\langle \psi' + \varphi, \mu \rangle = 0. \tag{2.67}$$

To obtain (2.67), we do an explicit computation: by (2.61)-(2.62), we have

$$\begin{aligned}\int_{\mathbb{R}} \varphi(z) d\mu(z) &= \int_{-\infty}^{0} \int_{-\infty}^{z} \bar{\psi}(r)\gamma(r) dr dz - \int_{0}^{+\infty} \int_{z}^{+\infty} \bar{\psi}(r)\gamma(r) dr dz \\ &= -\int_{\mathbb{R}} \bar{\psi}(r)\gamma(r) dr,\end{aligned}$$

by Fubini's Theorem. Since $-r\gamma(r) = \gamma'(r)$, integration by parts gives the result. Now we can differentiate the equation $L\varphi = \bar{\psi}$, to obtain

$$L\varphi' = \psi' + \varphi. \tag{2.68}$$

By (2.67) and (2.66), we deduce that

$$\|\varphi''\|_\infty \leq 2\|\psi'\|_\infty + 2\|\varphi\|_\infty. \tag{2.69}$$

To conclude, there remains to prove

$$\|\varphi\|_\infty \leq \|\psi'\|_\infty. \tag{2.70}$$

This is a consequence of the maximum principle for the elliptic equation (2.68). Indeed, assume $M = \sup_{z\in\mathbb{R}} \varphi(z) > 0$ (if $M \leq 0$, we have nothing to prove). Since φ tends to 0 at $\pm\infty$ by (2.65), M is reached at a point $z_M \in \mathbb{R}$. At this point, we have $\varphi'(z_M) = 0$ and $\varphi''(z_M) \leq 0$. In particular, $L\varphi'(z_M) \leq 0$. By (2.68), we deduce that $M = \varphi(z_M) \leq -\psi'(z_M) \leq \|\psi'\|_\infty$. A similar argument using $\inf_{z\in\mathbb{R}} \varphi(z)$ gives the result. □

2.11 Law of large numbers

Theorem 2.19 (Weak law of large numbers)**.** *Let $X_1, X_2, \ldots$ be some independent, identically distributed random variables with finite first moment $\mathbb{E}|X_1|$. Then, the average*

$$\bar{X}_N := \frac{X_1 + \cdots + X_N}{N}$$

is converging in probability to the constant $\mu = \mathbb{E}[X_1]$.

Theorem 2.20 (Strong law of large numbers)**.** *Let $X_1, X_2, \ldots$ be some independent, identically distributed random variables with finite first moment $\mathbb{E}|X_1|$. Then, the average*

$$\bar{X}_N := \frac{X_1 + \cdots + X_N}{N}$$

is converging almost-surely to the constant $\mu = \mathbb{E}[X_1]$.

Proof of Theorem 2.19. The mean value of $\bar{X}_N$ is

$$\mathbb{E}\left[\bar{X}_N\right] = \frac{\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_N]}{N} = \mu.$$

If the random variables X_n have finite second moment $\mathbb{E}|X_1|^2$, then, due to (2.28), the variance of $\bar{X}_N$ is

$$\mathbb{V}\mathrm{ar}(\bar{X}_N) = \frac{\mathbb{V}\mathrm{ar}(X_1) + \cdots + \mathbb{V}\mathrm{ar}(X_N)}{N^2} = \frac{\mathbb{V}\mathrm{ar}(X_1)}{N}. \tag{2.71}$$

Consequently $\bar{X}_N \to \mu$ in $L^2(\Omega)$ and thus (by the Markov inequality) in probability, when $N \to +\infty$. To treat the general case of $L^1(\Omega)$ random variables, we introduce the following truncates

$$T_R(X_n) = [X_n \wedge R] \vee (-R),$$

and the average

$$\bar{X}_{R,N} := \frac{T_R(X_1) + \cdots + T_R(X_N)}{N}.$$

Since $|T_R(X_n)| \leq R$, the variance of $T_R(X_n)$ is bounded by R^2, and the estimate (2.71) gives us the bound

$$\mathbb{E}|\bar{X}_{R,N} - \mu_R| \leq \mathbb{V}\mathrm{ar}(\bar{X}_{R,N})^{1/2} \leq \frac{R}{\sqrt{N}}, \tag{2.72}$$

where μ_R is the mean value of $T_R(X_n)$. The two averages $\bar{X}_{R,N}$ and $\bar{X}_N$ close to each other for large R. Indeed,

$$\mathbb{E}|\bar{X}_{R,N} - \bar{X}_N| \leq \frac{\mathbb{E}|X_1 - T_R(X_1)| + \cdots + \mathbb{E}|X_N - T_R(X_N)|}{N} = \mathbb{E}|X_1 - T_R(X_1)|,$$

and $\mathbb{E}|X_1 - T_R(X_1)| \to 0$ when $R \to +\infty$ by dominated convergence. Note that we have also

$$|\mu - \mu_R| \leq \mathbb{E}|X_1 - T_R(X_1)|$$

by the triangle inequality. Using (2.72), it follows that

$$\mathbb{E}|\bar{X}_N - \mu| \leq 2\mathbb{E}|X_1 - T_R(X_1)| + \frac{R}{\sqrt{N}}.$$

Choosing R large and then N large, we can make $\mathbb{E}|\bar{X}_N - \mu|$ arbitrary small. $\square$

3 Stochastic processes and the Brownian motion

Brownian motion is the type of motion observed by the botanist Robert Brown (1826-1827), specifically the motion of pollen particles, subject to collisions with the atom of the fluid in which they evolve. Brown noted the irregularity of the trajectories in particular. The mathematical theory of Brownian motion is due to Norbert Wiener (1920's). The mathematical model for Brownian motion is called a Wiener process, but also Brownian motion by extension.

Definition 3.1 (Stochastic process)**.** Let E be a separable Banach space, I a subset of $\mathbb{R}$ and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. An E-valued *stochastic process* $(X_t)_{t\in I}$ is a collection of random variables $X_t\colon \Omega \to E$ indexed by I.

If $I = \mathbb{N}$ or $\mathbb{Z}$, then $(X_t)_{t\in I}$ is a discrete-time process. We will work with continuous-time processes, for which $I = \mathbb{R}_+$ or $\mathbb{R}$ or $[0, T]$, $T > 0$. We will also use the notation $[0, T]$ for $\mathbb{R}_+$ when $T = +\infty$.

Definition 3.2 (Processes with independent increments)**.** Let E be a separable Banach space. A process $(X_t)_{t\in[0,T]}$ with values in E is said to have independent increments if, for all $n \in \mathbb{N}^*$, for all $0 \leq t_1 < \ldots < t_n \leq T$, the family $\{X_{t_{i+1}} - X_{t_i}; i = 1, \ldots, n-1\}$ of E-valued random variables is independent.

Definition 3.3 (Gaussian process)**.** A process $(X_t)_{t\in[0,T]}$ with values in $\mathbb{R}$ is said to be a Gaussian process if, for all $n \in \mathbb{N}^*$, for all $0 \leq t_1 < \ldots < t_n \leq T$, $(X_{t_i})_{i=1,\ldots,n}$ is a Gaussian vector in $\mathbb{R}^n$.

Definition 3.4 (Processes with continuous trajectories)**.** Let E be a separable Banach space. A process $(X_t)_{t\in[0,T]}$ with values in E is said to have continuous trajectories, if for all $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous from $[0, T]$ to E. If this is realized only almost surely (for ω in a set of full measure), then we say that (X_t) is almost surely continuous, or has almost surely continuous trajectories.

Similarly, one defines processes that are *càdlàg*: for all $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous from the right and has limit from the left (continue à droite, limite à gauche, *i.e.* càdlàg in french). We also speak of process with almost sure càdlàg trajectories. An example of càdlàg process is the Poisson process defined below. The trajectories of a process $(X_t)_{t\in[0,T]}$ may have more regularity than the C^0-regularity. Consider for example a process satisfying: there exists $\alpha \in (0, 1)$ such that, for $\mathbb{P}$-almost all $\omega \in \Omega$, there exists a constant $C(\omega) \geq 0$ such that

$$\|X_t(\omega) - X_s(\omega)\|_E \leq C(\omega)|t - s|^\alpha, \tag{3.1}$$

for all $t, s \in [0, T]$. Then we say that $(X_t)_{t\in[0,T]}$ has almost surely α-Hölder trajectories, or is almost-surely C^α. The Wiener process defined below has almost surely α-Hölder trajectories for all $\alpha < \frac{1}{2}$, *cf.* Corollary 3.5.

Definition 3.5 (Poisson process)**.** Let $\lambda > 0$. A Poisson process $(N_t)_{t\geq 0}$ with parameter λ is a process with values in $\mathbb{N}$ with *independent* increments satisfying $N_0 = 0$ almost-surely, $t \mapsto N_t$ is integer-valued, càdlàg and non-decreasing almost surely and

1. for $t, h \geq 0$, the law of $N(t+h) - N(t)$ is independent on t,
2. for all $t \geq 0$, $\mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)$,
3. for all $t \geq 0$, $\mathbb{P}(N_{t+h} - N_t > 1) = o(h)$.

One can show that, for $0 \leq s < t$, the increment $N_t - N_s$ has the Poisson distribution of parameter $\lambda(t-s)$:

$$\mathbb{P}(N_t - N_s = k) = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^k}{k!}. \tag{3.2}$$

One can also how that

$$N_t = \sum_{i=n}^{\infty} \mathbf{1}_{T_1+\cdots+T_n \geq t}, \tag{3.3}$$

where $T_1, \dots, T_n, \dots$ are i.i.d. random variable with exponential law of parameter λ. It is clear on Expression (3.3) that $t \mapsto N_t$ is integer-valued, càdlàg and non-decreasing almost surely. An example of process with almost sure continuous trajectories is the Wiener process.

Definition 3.6 (Wiener process)**.** A d-dimensional Wiener process is a process $(B_t)_{t\geq 0}$ with values in $\mathbb{R}^d$ such that: $B_0 = 0$ almost-surely, $(B_t)_{t\geq 0}$ has independent increments, and, for all $0 \leq s < t$, the increment $B_t - B_s$ follows the normal law $\mathcal{N}(0, (t-s)\mathrm{I}_d)$.

By standard properties of Gaussian $\mathbb{R}^n$-valued random variables ("Gaussian vectors"). The last two properties my be summed up in a single one: for all $n \in \mathbb{N}^*$, for all $0 \leq t_1 < \ldots < t_n$, setting $t_0 = 0$, the vector $(X_{t_i} - X_{t_{i-1}})_{i=1,\dots,n}$ of size nd is centred Gaussian with covariance the diagonal matrix with jj-entries $t_i - t_{i-1}$ for $j = id + 1, \dots, (i+1)d$, $0, \dots, n-1$. The almost-sure continuity of $t \mapsto B_t$ is often added in the definition, although it is a consequence (after modification) of the stated properties, see Corollary 3.5.

3.1 Law of a process

3.1.1 Cylindrical sets

Let E be a separable Banach space. A process $(X_t)_{t\in[0,T]}$ with values in E can be seen as a function

$$X \colon \Omega \to E^{[0,T]}, \tag{3.4}$$

where $E^{[0,T]}$ is the set of the applications $[0,T] \to E$. Let $\mathcal{F}_{\mathrm{cyl}}$ denote the cylindrical σ-algebra on $E^{[0,T]}$. This is the coarsest (minimal) σ-algebra that makes the projections

$$\pi_t \colon E^{[0,T]} \to E, \quad Y \mapsto Y_t$$

measurable. It is called cylindrical because it is generated by the *cylindrical sets*, which are subsets of $E^{[0,T]}$ of the form

$$D = \pi_{t_1}^{-1}(B_1) \bigcap \cdots \bigcap \pi_{t_n}^{-1}(B_n) = \left\{ Y \in E^{[0,T]}; Y_{t_1} \in B_1, \ldots, Y_{t_n} \in B_n \right\}, \tag{3.5}$$

where $t_1, \ldots, t_n \in [0, T]$ for a given $n \in \mathbb{N}^*$, and $B_1, \ldots, B_n$ are Borel subsets of E. Roughly speaking, in (3.5), D is the product of $B_1 \times \cdots \times B_n$ with the whole space $\prod_{t \neq t_j} E$. This why we speak of cylinder set. We have

$$X^{-1}(D) = \bigcap_{j=1}^{n} X_{t_j}^{-1}(B_j) \in \mathcal{F},$$

hence $X \colon (\Omega, \mathcal{F}) \to (E^{[0,T]}, \mathcal{F}_{\mathrm{cyl}})$ is a *random variable*.

Definition 3.7 (Law of a stochastic process)**.** Let E be a separable Banach space. The law of an E-valued stochastic process $(X_t)_{t \in [0,T]}$ is the probability measure μ_X on $(E^{[0,T]}, \mathcal{F}_{\mathrm{cyl}})$ induced by the map X in (3.4).

Remark 3.1*.* The σ-algebra $\mathcal{F}_{\mathrm{cyl}}$ being generated by the cylindrical sets, the law of X is characterized by the data

$$\mathbb{P}(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n),$$

which are called the *finite-dimensional distributions* of X (see Section 3.1.2).

We can be more specific on $\mathcal{F}_{\mathrm{cyl}}$. Each cylindrical set in (3.5) is of the form

$$\left\{ Y \in E^{[0,T]}; (Y_t)_{t \in J} \in B \right\}, \tag{3.6}$$

where J is a *countable* (since finite) subset of $[0, T]$ and B an element of the product σ-algebra $\Pi_{t \in J} \mathcal{B}(E_t)$, where $E_t = E$ for all t (this latter is the cylindrical σ-algebra for E^J). The collection of sets of the form (3.6) is precisely $\mathcal{F}_{\mathrm{cyl}}$.

Lemma 3.1 (Countably generated sets)**.** *The cylindrical σ-algebra $\mathcal{F}_{\mathrm{cyl}}$ is the collections of sets of the form* (3.6)*, for $J \subset [0, T]$ countable and B in the cylindrical σ-algebra of E^J.*

Proof of Lemma 3.1*.* Let us call $\mathcal{F}_\circ$ the collection of sets of the form (3.6), for $J \subset [0, T]$ countable and B in the cylindrical σ-algebra of E^J. The countable union of countable sets being countable, $\mathcal{F}_\circ$ is stable by countable union. Clearly it contains the empty set and is stable when taking the complementary since

$$\left\{ Y \in E^{[0,T]}; (Y_t)_{t \in J} \in B \right\}^c = \bigcup_{t \in J} \pi_t^{-1}(C_t), \quad C_t = (\pi_t(B))^c \in \mathcal{B}(E).$$

Therefore, $\mathcal{F}_\circ$ is a σ-algebra. Since $\mathcal{F}_\circ$ contains cylindrical sets (case J finite in (3.6)), $\mathcal{F}_\circ = \mathcal{F}_{\mathrm{cyl}}$. □

A corollary of this characterization of $\mathcal{F}_{\mathrm{cyl}}$ is that a lot of sets described in terms of an uncountable set of values X_t of the process $(X_t)_{t\in[0,T]}$ are *not measurable*, *i.e* not in $\mathcal{F}_{\mathrm{cyl}}$ (see the following exercise). This is due to the fact that $[0,T]$ is uncountable. For processes indexed by countable sets (discrete time processes), these problems of non-measurable sets do not appear.

Exercise 3.8. Show that the following sets are not in $\mathcal{F}_{\mathrm{cyl}}$:

1. $A_1 = \{X \equiv 0\} = \bigcap_{t\in[0,T]} \pi_t^{-1}(\{0\})$,

2. $A_2 = \{t \mapsto X_t \text{ is continuous}\}$.

The solution to Exercise 3.8 is here.

Now, assume that $(X_t)_{t\in[0,T]}$ is a process with almost-sure continuous trajectories. Then we would like to say that, instead of (3.4), we have

$$X\colon \Omega \to C([0,T];E), \tag{3.7}$$

In that case, the sets A_1 and A_2 in Exercise 3.8 *are* measurable.

Exercise 3.9. Let

$$\mathcal{F}_{\mathrm{cts}} = \mathcal{F}_{\mathrm{cyl}} \cap C([0,T];E).$$

Show that the σ-algebra $\mathcal{F}_{\mathrm{cts}}$ coincides with the Borel σ-algebra on $C([0,T];E)$, the topology on $C([0,T];E)$ being the topology of Banach space with norm

$$X \mapsto \sup_{t\in[0,T]} \|X(t)\|_E.$$

Then show that the sets A_1 and A_2 in Exercise 3.8 are measurable.
The solution to Exercise 3.9 is here.

Actually, starting from (3.4), we have (3.7) indeed only if we first redefine X on $\Omega\setminus\Omega_{\mathrm{cts}}$ where Ω_{cts} is the set of ω such that $t\mapsto X_t(\omega)$ is continuous. However, it is not ensured that $\Omega_{\mathrm{cts}}(=X^{-1}(A_2)$ with the notation of Exercise 3.8) is measurable. A correct procedure is the following one (we modify not only Ω, but $\mathbb{P}$ also [RY99]). Define the probability measure Q on $\mathcal{F}_{\mathrm{cts}}$ by

$$Q(A) = \mathbb{P}(X \in \tilde{A}), \quad A = \tilde{A} \cap C([0,T];E), \quad \tilde{A} \in \mathcal{F}_{\mathrm{cyl}}. \tag{3.8}$$

for all $A \in \mathcal{F}_{\mathrm{cts}}$. By definition, each $A \in \mathcal{F}_{\mathrm{cts}}$ can be written as in (3.8). If two decompositions

$$A = \tilde{A}_1 \cap C([0,T];E) = \tilde{A}_2 \cap C([0,T];E)$$

are possible, then the definition of $Q(A)$ is unambiguous since $\mathbb{P}(X \in \tilde{A}_1) = \mathbb{P}(X \in \tilde{A}_2)$. Indeed, by hypothesis, there exists a measurable subset G of Ω of full measure such that: $\omega \in G$ implies that $t \mapsto X_t(\omega)$ is continuous (*i.e.* $G \subset \Omega_{\mathrm{cts}}$). If $\omega \in X^{-1}(\tilde{A}_1)\cap G$, then

$$X(\omega) \in \tilde{A}_1 \cap C([0,T];E) = \tilde{A}_2 \cap C([0,T];E),$$

hence $X^{-1}(\tilde{A}_1) \cap G \subset X^{-1}(\tilde{A}_2) \cap G$. It follows that

$$\mathbb{P}(X \in \tilde{A}_1) = \mathbb{P}(X^{-1}(\tilde{A}_1) \cap G) \leq \mathbb{P}(X^{-1}(\tilde{A}_1) \cap G) = \mathbb{P}(X \in \tilde{A}_2).$$

By symmetry of $\tilde{A}_1$ and $\tilde{A}_2$, we obtain the result. We consider then the canonical process

$$Y_t \colon C([0,T];E) \to \mathbb{R}, \quad Y_t(\omega) = \omega(t).$$

The law of Y on $(C([0,T];E), \mathcal{F}_{\mathrm{cts}}, Q)$ is the same as X (*cf.* Remark 3.1), thus considering X or Y is equivalent, and Y has the desired path-space $C([0,T];E)$.

3.1.2 Finite-dimensional distributions

Let us introduce the following notation: if I is a subset of $[0,T]$ we denote by π_I the projection $E^{[0,T]} \to E^I$ which maps $(Y_t)_{t\in[0,T]}$ to $(Y_t)_{t\in I}$. We have then a probability measure P_I on E^I defined as $P_I = [\pi_I]_* \mu_X$. If I is finite, say $I = \{t_1, \ldots, t_n\}$, then

$$P_I(B) = \mathbb{P}((X_{t_1}, \ldots, X_{t_n}) \in B)$$

for all B in the product σ-algebra $\Pi_{t\in I}\mathcal{B}(E)$. The probability measures P_I, for I finite are called the finite-dimensional distributions of $(X_t)_{t\in[0,T]}$. They satisfy the consistency relation

$$P_J = [\pi_{J\leftarrow I}]_* P_I \tag{3.9}$$

for all $J \subset I \subset [0,T]$ with I finite, where $\pi_{J\leftarrow I}$ is the projection $E^I \to E^J$. The Kolmogorov extension theorem asserts that, to any collection of finite distribution satisfying the consistency relation corresponds a unique probability measure μ on $(E^{[0,T]}, \mathcal{F}_{\mathrm{cyl}})$ such that $P_i = [\pi_I]_* \mu_X$ for all finite $I \subset [0,T]$, see [Tao11, Theorem 2.4.3] for a more precise statement. We will not use the Kolmogorov extension theorem, but we will mention some of its corollaries when it is relevant.

Example 3.2 (Wiener process). Let $t_1, \ldots, t_n \in \mathbb{R}_+$. Assume the times are ordered: $t_1 < \cdots < t_n$. Let $\theta_n \colon \mathbb{R}^{nd} \to \mathbb{R}^{nd}$ denote the map

$$(w_1, \ldots, w_n) \mapsto (w_1, w_2 - w_1, \ldots, w_n - w_{n-1}), \quad w_1, \ldots, w_n \in \mathbb{R}^d. \tag{3.10}$$

The application θ_n is a linear isomorphism of inverse

$$\theta_n^{-1} \colon (z_1, \ldots, z_n) \mapsto (z_1, z_2 + z_1, \ldots, z_n + z_{n-1} + \cdots + z_1). \tag{3.11}$$

The Wiener process $(W_t)_{t\in\mathbb{R}_+}$ is such that $\theta_n((W_{t_i})_{1,n})$ is centred Gaussian with covariance the diagonal matrix $\Gamma_{(t)}$ with jj-entries $t_i - t_{i-1}$ for $j = id+1, \ldots, (i+1)d$, $i = 0, \ldots, n-1$ (with the convention $t_0 = 0$), *cf.* the comment after Definition 3.6. The finite-dimensional distribution $[\mathbb{P}_W]_{t_1,\ldots,t_k}$ is therefore given as $[\theta_n^{-1}]_* \tilde{P}_{t_1,\ldots,t_n}$, where $\tilde{P}_{t_1,\ldots,t_n}$ is the $\mathcal{N}(0, \Gamma_{(t)})$ law. It is clear that the consistency relation (3.9) is satisfied. The Kolmogorov extension theorem gives therefore the existence of a one-dimensional Wiener process. The existence of a one-dimensional Wiener process will also be established by the Donsker theorem, see Section 3.4.

Example 3.3 (Poisson process). Let $(N_t)_{t\geq 0}$ be a Poisson process of parameter λ. Let $t_1, \ldots, t_n \in \mathbb{R}_+$. Assume again that the times are ordered: $t_1 < \cdots < t_n$. Recall (3.2): the increments $N_{t_{i+1}} - N_{t_i}$ have the Poisson distribution of parameter $\lambda(t_{i+1} - t_i)$. By independence, this gives the probabilities

$$\mathbb{P}(N_{t_2} - N_{t_1} \in B_1, \ldots, N_{t_n} - N_{t_{n-1}} \in B_n),$$

for $B_1, \ldots, B_n \subset \mathbb{N}$ from which we deduce, using (3.11), the finite-dimensional distributions $P_{t_1,\ldots,t_n}$. Again, we check that we can apply the Kolmogorov extension theorem, to obtain the existence of a Poisson process with parameter λ.

The concept of *stationary* process is based on the finite-dimensional distribution.

Definition 3.10 (Stationary process)**.** Let E be a separable Banach space. An E-valued stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *stationary* if, for all $\sigma \geq 0$, for all finite $I \subset \mathbb{R}_+$, $P_I = P_{\sigma+I}$:

$$\mathbb{P}(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n) = \mathbb{P}(X_{\sigma+t_1} \in B_1, \ldots, X_{\sigma+t_n} \in B_n) \tag{3.12}$$

for all $t_1, \ldots, t_n \in \mathbb{R}_+$, $B_1, \ldots, B_n \in \mathcal{B}(E)$.

We will also consider later stationary process $(X_t)_{t\in\mathbb{R}}$ indexed by $\mathbb{R}$, in which case (3.12) is satisfied for all $t_1, \ldots, t_n \in \mathbb{R}$ and all $\sigma \in \mathbb{R}$.

Exercise 3.11. This exercise is about stationary processes. The questions are independent.

1. Are the Wiener process and the Poisson process of parameter λ stationary processes?

2. Show that if $(X_t)_{t\in\mathbb{R}_+}$ is a stationary process, then the law of X_t is constant in time.

3. Let $X_0, X_1, \ldots$ be the sequence of random variables on $\mathbb{R}$ defined as follows: X_0 is chosen at random, according to a law μ_0, then, X_N being known, a random variable Z_{N+1} taking the values $+1$ or -1 with equi-probability is drawn independently on $X_0, \ldots, X_N$ and X_{N+1} is given by

$$X_{N+1} = \frac{1}{2} X_N + Z_{N+1}. \tag{3.13}$$

(This time-discrete process was already considered in Exercise 2.8). We assume that μ_0 is the uniform measure of the interval $[-2, 2]$:

$$\int_{\mathbb{R}} \varphi(x) d\mu_0(x) = \frac{1}{4} \int_{-2}^{2} \varphi(x) dx.$$

   (a) Show that the law of X_n is independent on n.
   (b) Show that $(X_n)_{n\in\mathbb{N}}$ is stationary.

The solution to Exercise 3.11 is here.

3.1.3 Equality of processes

In all this section, E is a separable Banach space. The Borel σ-algebra is denoted $\mathcal{B}(E)$. We have seen in the previous section 3.1.2 that the law of an E-valued processes $(X_t)_{t\in[0,T]}$ is characterized by the collection of the finite-dimensional distributions. Therefore, two E-valued processes $(X_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ have the *same law* if, and only if,

$$\mathbb{P}(X_{t_1}\in B_1,\dots,X_{t_n}\in B_n)=\mathbb{P}(Y_{t_1}\in B_1,\dots,Y_{t_n}\in B_n) \tag{3.14}$$

for all $t_1,\dots,t_n\in[0,T]$, $B_1,\dots,B_n\in\mathcal{B}(E)$. In that case, we say that the two processes are *equivalent*, or that $(Y_t)_{t\in[0,T]}$ is a *version* of $(X_t)_{t\in[0,T]}$. The equality of finite-dimensional distributions (3.14) is satisfied in particular if

$$\text{for all } t\in[0,T],\quad X_t=Y_t \text{ almost-surely,} \tag{3.15}$$

which means, according to the following definition, that $(Y_t)_{t\in[0,T]}$ is a modification of $(X_t)_{t\in[0,T]}$.

Definition 3.12 (Modification of a process)**.** Let $(X_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ be two E-valued processes. If (3.15) is satisfied, then we say that $(Y_t)_{t\in[0,T]}$ is a *modification* of $(X_t)_{t\in[0,T]}$.

Modifications of processes do not affect their statistical properties thus, but modify their regularity or measurability properties, see Theorem 3.2 and Theorem 3.3 below.

3.2 Elementary properties of processes

Our main aim, in this section, is to prove the following result.

Theorem 3.2 (Kolmogorov's continuity theorem)**.** *Let E be a separable Banach space. Let $(X_t)_{t\in[0,T]}$ be a process with values in E which satisfies*

$$\mathbb{E}\|X(t)-X(s)\|_E^p\leq C|t-s|^{1+\delta}, \tag{3.16}$$

for all $s,t\in[0,T]$, where $p>1$, $\delta>0$ and $C\geq 0$ are some given constant. Then $(X_t)_{t\in[0,T]}$ has a modification which has α-Hölder trajectories for any $\alpha<\frac{\delta}{p}$.

Proof of Theorem 3.2. We give the proof of [DPZ92, p. 73], based on the Sobolev embedding Theorem: if $r\geq 1$, $\sigma>0$, $\sigma r>1$, then

$$W^{\sigma,r}(0,T;E)\hookrightarrow C^\mu([0,T];E) \tag{3.17}$$

for all $\mu\in(0,\sigma-\frac{1}{r})$. In (3.17), $W^{\sigma,r}(0,T;E)$ is the space of functions u in $L^r(0,T;E)$ with finite norm

$$\|u\|_{W^{\sigma,r}(0,T;E)}=\|u\|_{L^r(0,T;E)}+\left[\int_0^T\int_0^T\frac{\|u(t)-u(s)\|_E^r}{|t-s|^{1+\sigma r}}dsdt\right]^{\frac{1}{r}}.$$

Note that (3.17) has to be understood as follows: every $u \in W^{\sigma,r}(0,T;E)$ has a representative $\tilde{u} \in C^\mu([0,T];E)$ with

$$\|\tilde{u}\|_{C^\mu([0,T];E)} := \sup_{t\in[0,T]} \|\tilde{u}(t)\|_E + \sup_{t,s\in[0,T]} \frac{\|\tilde{u}(t)-\tilde{u}(s)\|_E}{|t-s|^\mu} \leq C\|u\|_{W^{\sigma,r}(0,T;E)}, \tag{3.18}$$

where C is a constant depending on r, σ, μ. Note furthermore (we will use it later) that the most natural candidate for the representative $\tilde{u}$ is the function $[u]$ defined by

$$[u](t) = \begin{cases} \lim_{n\to+\infty} \frac{1}{|J_n(t)|} \int_{J_n(t)} u(s)ds, & \text{if the limit exists,} \\ 0 \quad \text{otherwise,} \end{cases} \tag{3.19}$$

where $J_n(t)$ is the interval $J_n(t) := [t-n^{-1}, t+n^{-1}] \cap [0,T]$ and $|J_n(t)|$ its length. Indeed, by the Lebesgue differentiation theorem [Tao11, Theorem 1.6.12], for almost every $t \in [0,T]$, the limit in (3.19) exists and $[u](t) = u(t)$. Besides $u = [u]$ if u is continuous. To prove the theorem, let us assume

$$\mathbb{E}\|X(0)\|_E^p \leq C. \tag{3.20}$$

Actually, we may as well suppose that $X(0) = 0$. Indeed, if not, we prove the result for $\tilde{X}(t) = X(t) - X(0)$ (this process satisfies (3.16) also). If $(\tilde{Y}_t)_{[0,T]}$ is a modification of $(\tilde{X}_t)_{[0,T]}$ with α-Hölder trajectories, then $(X(0) + \tilde{Y}_t)_{[0,T]}$ is a modification of $(X_t)_{[0,T]}$ with α-Hölder trajectories,

Let $\sigma < \frac{1+\delta}{p}$. Let us apply the Fubini theorem to the functions

$$(t,\omega) \mapsto \|X(t,\omega)\|_E^p, \quad (t,s,\omega) \mapsto \frac{\|X(t,\omega)-X(s,\omega)\|_E^p}{|t-s|^{1+\sigma p}}. \tag{3.21}$$

Since $[0,T]\times\Omega$ has finite measure under $dt \times d\mathbb{P}$ and $[0,T]\times[0,T]\times\Omega$ has finite measure under $dt\times ds\times d\mathbb{P}$, this is licit if the functions in (3.21) are measurable. This point is not obvious, and actually requires to consider a modification of the process $(X_t)_{t\in[0,T]}$. We postpone the discussion of that fact to the end of the proof. For the moment, interchanging integrals, we obtain thus

$$\mathbb{E}\|X\|_{W^{\sigma,p}(0,T;E)}^p = \int_0^T \mathbb{E}\|X(t)\|_E^p dt + \int_0^T \int_0^T \frac{\mathbb{E}\|X(t)-X(s)\|_E^p}{|t-s|^{1+\sigma p}} dsdt.$$

By (3.16)-(3.20), we have

$$\mathbb{E}\|X(t)\|_E^p \leq 2^p \left(\mathbb{E}\|X(t)-X(0)\|_E^p + \mathbb{E}\|X(0)\|_E^p\right) \leq C(T),$$

which gives

$$\mathbb{E}\|X\|_{W^{\sigma,p}(0,T;E)}^p \leq TC(T) + C\int_0^T \int_0^T |t-s|^{\delta-\sigma p} dsdt < +\infty, \tag{3.22}$$

thanks to the condition $\sigma < \frac{1+\delta}{p}$. Let

$$\Omega_\sigma = \{\|X\|^p_{W^{\sigma,p}(0,T;E)} < +\infty\}.$$

Then Ω_σ is measurable (a consequence of the fact that the functions in (3.21) are measurable) and of full measure, $\mathbb{P}(\Omega_\sigma) = 1$. Indeed

$$\Omega_\sigma^c = \bigcap_{n\in\mathbb{N}^*} A_n, \quad A_R := \left\{\|X\|^p_{W^{\sigma,p}(0,T;E)} > R\right\}, \tag{3.23}$$

a decreasing intersection, and by the Markov inequality,

$$\mathbb{P}(A_R) \leq \frac{1}{R}\mathbb{E}\|X\|^p_{W^{\sigma,p}(0,T;E)}. \tag{3.24}$$

Let (σ_m) be an increasing sequence of positive reals converging to $\frac{1+\delta}{p}$. We set $\tilde{\Omega} = \cap_{m\in\mathbb{N}}\Omega_{\sigma_m}$. Then $\mathbb{P}(\tilde{\Omega}) = 1$. We also set

$$Y(t,\omega) = \begin{cases} [X(\cdot,\omega)](t), & \omega \in \tilde{\Omega}, \\ 0, & \omega \notin \tilde{\Omega}, \end{cases}$$

where $[u]$ is defined in (3.19). Then Y_t is a random variable for all $t \in [0,T]$ and for all $\omega \in \tilde{\Omega}$, for all $\mu < \frac{\delta}{p}$, $t \mapsto Y(t,\omega)$ is in $C^\mu([0,T])$ (choose $\sigma_m < \frac{1+\delta}{p}$ such that $\mu < \sigma_m - \frac{1}{p}$ and use the fact that $\tilde{\Omega} \subset \Omega_{\sigma_m}$). Therefore $(Y_t)_{t\in[0,T]}$ has almost-surely μ-Hölder trajectories. We know also that, for $\mathbb{P}$-almost all ω (the ω's in $\tilde{\Omega}$), for a.e. $t \in [0,T]$, $Y_t(\omega) = X_t(\omega)$. To prove that $(Y_t)_{t\in[0,T]}$ is a modification of $(X_t)_{t\in[0,T]}$, we have to invert the order of ω and t in this statement, and to have a result for all t. Let us fix $t \in [0,T]$ therefore. We have the bound

$$\mathbb{E}\left\|X(t) - \frac{1}{|J_n(t)|}\int_{J_n(t)} X(s,\omega)ds\right\|_E \leq \frac{1}{|J_n(t)|}\int_{J_n(t)} \left[\mathbb{E}\|X(t) - X(s)\|_E^p\right]^{1/p} ds$$

by the Hölder inequality. We deduce from (3.16) that

$$\mathbb{E}\left\|X(t) - \frac{1}{|J_n(t)|}\int_{J_n(t)} X(s,\omega)ds\right\|_E \leq \frac{C^{1/p}}{n^{(1+\delta)/p}}.$$

It follows, up to a subsequence, that

$$\frac{1}{|J_n(t)|}\int_{J_n(t)} X(s,\omega)ds \to X(t,\omega)$$

for all $\omega \in \hat{\Omega}$, where $\mathbb{P}(\hat{\Omega}) = 1$. If $\omega \in \hat{\Omega} \cap \tilde{\Omega}$ (a set of full measure in Ω), we have also

$$\frac{1}{|J_n(t)|}\int_{J_n(t)} X(s,\omega)ds \to Y(t,\omega),$$

and, consequently, $Y(t,\omega) = X(t,\omega)$. This proves that $X_t = Y_t$ almost-surely.

There remains to prove the fact that we left aside, that, under (3.16), $(X_t)_{t\in[0,T]}$ has a modification $(\tilde{X}_t)_{t\in[0,T]}$ such that $(t,\omega) \mapsto \tilde{X}(t,\omega)$ is $\mathcal{B}([0,T]) \times \mathcal{F}$-measurable (this is sufficient for the functions in (3.21), with $\tilde{X}$ in place of X, to be measurable). Let us first observe that (3.16) implies that $(X_t)_{t\in[0,T]}$ is *stochastically continuous*, which means continuous for the convergence in probability: for all $t \in [0,T]$, for all $\varepsilon > 0$,

$$\lim_{s\to t} \mathbb{P}(\|X(t) - X(s)\|_E > \varepsilon) = 0. \tag{3.25}$$

Indeed, by the Markov inequality and (3.16), we have

$$\mathbb{P}(\|X(t) - X(s)\|_E > \varepsilon) \le \varepsilon^{-p} C|t-s|^{1+\delta} \to 0 \text{ when } s \to t.$$

Then the result follows from Theorem 3.3 below. $\square$

Remark 3.4. Note that, thanks to (3.18) and to the bounds established in the proof of the Kolmogorov's continuity criterion, we have obtained the estimate

$$\mathbb{E}\|Y\|_{C^\alpha([0,T];E)} \le \mathbb{E}\|X_0\|_E + \tilde{C}, \tag{3.26}$$

for the modification $(Y_t)_{t\in[0,T]}$ of $(X_t)_{t\in[0,T]}$. In (3.26), the constant $\tilde{C}$ depends on T, E, p, δ and on the constant C in (3.16) only.

Definition 3.13 (Stochastically continuous process)**.** Let $(X_t)_{t\in[0,T]}$ be an E-valued process. It is said to be *stochastically continuous* at $t_* \in [0,T]$ if (X_t) is converging to t_* in probability when $t \to t_*$. It is said to be stochastically continuous without specific mention of the point if it is stochastically continuous at every $t \in [0,T]$.

Theorem 3.3 (Measurable modification)**.** *Let $(X_t)_{t\in[0,T]}$ be an E-valued, stochastically continuous process. Then $(X_t)_{t\in[0,T]}$ has a modification $(\tilde{X}_t)_{t\in[0,T]}$ such that $(t,\omega) \mapsto \tilde{X}(t,\omega)$ is $\mathcal{B}([0,T]) \times \mathcal{F}$-measurable.*

The following exercises introduce the tools used in the proof of Theorem 3.3.

Exercise 3.14. Prove the Borel-Cantelli lemma:

Lemma 3.4 (Borel-Cantelli)**.** *Let (A_n) be a sequence of events such that the series $\sum \mathbb{P}(A_n)$ is convergent. Then, almost-surely, a finite number of the A_n's is realized.*

The solution to Exercise 3.14 is here.

Exercise 3.15. Let (X_n) be a sequence of random variable which converges *rapidly* to 0 in probability, in the sense that, for every $\delta > 0$, the series $\sum \mathbb{P}(\|X_n\|_E > \delta)$ is convergent[3]. Show that (X_n) converges to 0 almost-surely.

The solution to Exercise 3.15 is here.

[3]Recall that convergence in probability only requires $\lim_{n\to+\infty} \mathbb{P}(\|X_n\|_E > \delta) = 0$ for every $\delta > 0$

Proof of Theorem 3.3. We will construct a sequence of $\mathcal{B}([0,T]) \times \mathcal{F}$-measurable functions which converges $dt \times d\mathbb{P}$-almost everywhere to a function Y such that $(Y_t)_{t\in[0,T]}$ is a modification of $(X_t)_{t\in[0,T]}$. First, we observe that stochastic continuity on $[0,T]$ implies uniform stochastic continuity. Indeed, if $\varepsilon, \delta > 0$, then, for every $t \in [0,T]$, there is an open neighbourhood J_t of t such that $s \in J_t$ implies $\mathbb{P}(\|X(t) - X(s)\|_E > \delta) < \varepsilon$. Covering the compact $[0,T]$ by a finite number of the intervals J_t, $t \in [0,T]$, we deduce that there exists $\eta > 0$ such that $t, s \in [0,T]$, $|t-s| < \eta$, imply $\mathbb{P}(\|X(t) - X(s)\|_E > \delta) < \varepsilon$. Let $(\delta_k), (\varepsilon_k) \downarrow 0$ and let η_k be the associated modulus of uniform stochastic continuity. We do a partition

$$[0,T] = \bigcup_{j=1}^{N_k} I_j^k$$

in intervals of length$< \eta_k$, pick up $t_j^k \in I_j^k$ and define

$$X_k(t,\omega) = \sum_{j=1}^{N_k} \mathbf{1}_{I_j^k}(t) X_{t_j^k}.$$

Then X_k is $\mathcal{B}([0,T]) \times \mathcal{F}$-measurable since, for $B \in \mathcal{B}(E)$, $X_k^{-1}(B)$ is the union over j of the measurable rectangles $I_j^k \times X_{t_j^k}^{-1}(B)$. Let A_{cv} denote the set of $(t,\omega) \in [0,T] \times \Omega$ such that the sequence $(X_k(t,\omega))$ is converging in E. Then A_{cv} is in $\mathcal{B}([0,T]) \times \mathcal{F}$ (this is a classical fact: we use a Cauchy criterion for the characterization of the convergence). Set

$$Y(t,\omega) = \begin{cases} \lim\limits_{k\to+\infty} X_k(t,\omega) \text{ if } (t,\omega) \in A_{\mathrm{cv}}, \\ 0 \text{ otherwise.} \end{cases}$$

Then Y is $\mathcal{B}([0,T]) \times \mathcal{F}$-measurable. To conclude, assume that the series $\sum \varepsilon_k$ is convergent. Let $t \in [0,T]$. For $k \geq 1$, we have $\|X(t) - X_k(t)\|_E = \|X(t) - X(t_j^k)\|_E$, where $j \in \{1, \ldots, N_k\}$ is the index such that $t \in I_j^k$. Since $|t - t_j^k| < \eta_k$, we have

$$\mathbb{P}(\|X(t) - X_k(t)\|_E > \delta_k) = \mathbb{P}(\|X(t) - X(t_j^k)\|_E > \delta_k) < \varepsilon_k.$$

From the Borel-Cantelli lemma, we deduce that there is a measurable set $\tilde{\Omega}_t \subset \Omega$ of probability 1 such that, for $\omega \in \tilde{\Omega}_t$, $\|X(t,\omega) - X_k(t,\omega)\|_E > \delta_k$ occurs only a finite number of time. Then we have $X_k(t,\omega) \to X(t,\omega)$. Consequently, $X(t,\omega) = Y(t,\omega)$ for all $\omega \in \tilde{\Omega}_t$. This concludes the proof. ☐

Corollary 3.5 (Hölder continuity of the Brownian motion). *Let $(B_t)_{t\geq 0}$ be a Wiener process. There is a modification of $(B_t)_{t\geq 0}$ which has α-Hölder continuous trajectories for all $\alpha < \frac{1}{2}$.*

Proof of Corollary 3.5. Let $p \in \mathbb{N}^*$, and $0 \leq s \leq t$. Since $B_t - B_s$ is normally distributed, with mean 0 and variance $t-s$, we have

$$\mathbb{E}|B_t - B_s|^p = \frac{1}{(2\pi(t-s))^{d/2}} \int_{\mathbb{R}} |x|^p e^{-\frac{|x|^2}{(t-s)}} dx = C_p (t-s)^{p/2}.$$

The last identity is obtained by the change of variable $x = (t-s)^{1/2}x'$. The constant C_p is the p-th moment of a $\mathcal{N}(0, \mathrm{I}_d)$ random variable. We choose $p > 2$ and let

$$\alpha_p := \frac{1}{p}\left(\frac{p}{2} - 1\right) = \frac{1}{2} - \frac{1}{p}.$$

We apply the Kolmogorov continuity theorem to obtain a modification of $(B_t)_{t\geq 0}$ which has α-Hölder continuous trajectories for all $\alpha < \alpha_p$. The modification may depend on p and we denote it by $(\tilde{B}_t^{(p)})_{t\geq 0}$. For each $t \geq 0$, $\tilde{B}_t^{(p)}$ coincides with B_t on a set $\Omega_t^{(p)}$ of probability 1. We define then $\tilde{B}_t = B_t$ on

$$\tilde{\Omega} := \bigcap_{p\in\mathbb{N}, p>2} \Omega_t^{(p)},$$

and $\tilde{B}_t = 0$ on the complementary of $\tilde{\Omega}$. The process $(\tilde{B}_t)_{t\geq 0}$ is a modification of $(B_t)_{t\geq 0}$ which has α-Hölder continuous trajectories for all $\alpha < \frac{1}{2}$. $\square$

Exercise 3.16. Admit the *Garsia - Rodemich - Rumsey inequality:* for all $r > 1$, $\sigma \in (r^{-1}, 1)$, there exists a constant $C_{\sigma,r}$, such that, for all continuous function $u\colon [0,T] \to \mathbb{R}$, one has

$$\|u(t) - u(s)\|_E^r \leq C_{\sigma,r}|t-s|^{\sigma r - 1} \int_0^T \int_0^T \frac{\|u(t') - u(s')\|_E^r}{|t'-s'|^{1+\sigma r}} ds' dt'. \tag{3.27}$$

Suppose that $(X_t)_{t\in[0,T]}$ satisfies the hypothesis of Theorem 3.2. Let $\alpha < \frac{\delta}{p}$. Show that there exists a modification $(\tilde{X}_t)_{t\in[0,T]}$ of $(X_t)_{t\in[0,T]}$ and a non-negative random variable ζ with moments of all orders less than p such that

$$\|\tilde{X}(t) - \tilde{X}(s)\|_E \leq \zeta |t-s|^{\alpha}, \tag{3.28}$$

almost-surely, for all $t, s \in [0,T]$.

Note: actually, the Garsia - Rodemich - Rumsey inequality is a little bit more general, see Section 1.3 of this course by S.R.S. Varadhan and [Bau14, Theorem 7.34] for example.

The solution to Exercise 3.16 is here.

3.3 The Wiener measure

We have seen in (2.6), (2.35) that is it often natural and/or useful to modify the quadruplet $(\Omega, \mathcal{F}, \mathbb{P}, X)$ into $(E, \mathcal{B}(E), \mu_X, \mathrm{Id}_E)$. Consider now an E-valued process $(X_t)_{t\in[0,T]}$ with continuous trajectories. Here E is a separable Banach space. We have seen in Section 3.1 (see (3.7) in particular) that we may see X as element of $C([0,T]; E)$. Then $F := C([0,T]; E)$ is the *path space*. Endowed with the norm $\sup_{t\in[0,T]} \|\omega(t)\|_E$, it is a separable Banach space. Recall that $\mathcal{B}(F)$ coincides with the trace on F of the cylindrical σ-algebra $\mathcal{F}_{\mathrm{cyl}}$ (Exercise 3.9). Consider the probability space

$$(F, \mathcal{B}(F), \mu_X).$$

The family of evaluations $(e_t)_{t\in[0,T]}$, where

$$e_t \colon F \ni \omega \mapsto \omega(t) \in \mathbb{R}$$

is the *canonical* process on F. It has the law μ_X.

The Wiener measure (often denoted $\mathbb{P}_W$) is the probability law on $C([0,T];\mathbb{R}^d)$ determined by the Wiener process, *i.e.* μ_B. It gives a way do draw continuous curves in $\mathbb{R}^d$, those latter being described thanks to a continuous parametrization $\omega\colon [0,T] \to \mathbb{R}^d$. The Wiener measure $\mathbb{P}_W$ is entirely characterized by the finite-dimensional distributions, described in Example 3.2.

3.4 The Donsker Theorem

Consider the random walk defined in Example 2.2, which evolves from time $n = 0$ to a final time N. We see it as a graph (with the interval $[0, N]$ in abscissa and the line $\mathbb{R}$ in ordinate) by using linear interpolation between the points (n, X_n). For the homogeneity of notations, we will use the notation S_n for X_n (S_n is the sum $Z_1 + \cdots + Z_n$ thus). Rescale this graph by a factor N^{-1} in abscissa and $N^{-1/2}$ in ordinate. This gives us a process $(\xi^N(t))$ defined, for $t \in [0, 1]$, by

$$\xi_N(t) = \frac{t_{i+1} - t}{t_{i+1} - t_i} \frac{S_i}{\sqrt{N}} + \frac{t - t_i}{t_{i+1} - t_i} \frac{S_{i+1}}{\sqrt{N}}, \quad t_i \leq t < t_{i+1}, \tag{3.29}$$

where $t_i = \frac{i}{N}$. More generally, we will consider (3.29) where S_n is the sum

$$S_n = \frac{Z_1 + \cdots + Z_n}{\sigma}$$

and $Z_1, Z_2, \ldots$ are independent identically distributed random variables, centred, with variance σ^2 and a finite forth moment: $\mathbb{E}|Z|^4 < +\infty$.

3.4.1 Finite-dimensional distributions

Proposition 3.6. *For all $t_1, \ldots, t_k \in [0, 1]$, for all $A_1, \ldots, A_k$ Borel subsets of $\mathbb{R}$, we have*

$$\lim_{N\to+\infty} \mathbb{P}(\xi_N(t_1) \in A_1, \ldots, \xi_N(t_k) \in A_k) = [\mathbb{P}_W]_{t_1,\ldots,t_k}(A_1, \ldots, A_k), \tag{3.30}$$

where $[\mathbb{P}_W]_{t_1,\ldots,t_k}$ is the finite-dimensional distribution of a one-dimensional Wiener process, introduced in Example 3.2.

Proof of Proposition 3.6. We will prove (3.30) simply for $k = 2$. For a general k the proof is similar. For $k = 1$ first, and $t \in (0, 1]$, we have (3.29) for $i = [Nt]$, where $[Nt]$ is the integer part of Nt. We rewrite (3.29) as

$$\xi_N(t) = \frac{S_i}{\sqrt{N}} + \varepsilon_i(t), \quad \varepsilon_i(t) := \frac{t - t_i}{t_{i+1} - t_i} \frac{S_{i+1} - S_i}{\sqrt{N}}. \tag{3.31}$$

The remainder $\varepsilon_i(t)$ converges to 0 in $L^2(\Omega)$ when $N \to +\infty$ since

$$|\varepsilon_i(t)| \leq \frac{|Z_{i+1}|}{\sigma\sqrt{N}},$$

and thus $\mathbb{E}|\varepsilon_i(t)|^2 \leq N^{-1}$. We have also

$$\frac{S_i}{\sqrt{N}} = \sqrt{\frac{[Nt]}{N}} \frac{S_{[Nt]}}{\sqrt{[Nt]}}.$$

The factor $\sqrt{\frac{[Nt]}{N}}$ tends to $\sqrt{t}$ when $N \to +\infty$. By the central limit theorem (*cf.* Theorem 2.17 and (2.59)), $\frac{S_{[Nt]}}{\sqrt{[Nt]}}$ is converging in law to the centred reduced $\mathcal{N}(0,1)$ law. We use Lemma 2.12 (and the scaling $\sqrt{t}\mathcal{N}(0,1) = \mathcal{N}(0,t)$) to conclude that $\xi_N(t)$ converges in law to the $\mathcal{N}(0,t)$ law, which is (3.30) for $k=1$. For $k=2$ now, let $0 < s < t \leq 1$ and let θ_2 be the function defined by (3.10). We have $\theta_2(\xi_N(s), \xi_N(t)) = (X_N, Y_N)$, where $X_N = \xi_n(s)$ is converging in law to $\mathcal{N}(0,s)$ as we saw, while

$$Y_N = \frac{Z_{[Ns]+1} + \cdots + Z_{[Nt]}}{\sigma\sqrt{N}} + \tilde{\varepsilon}_N, \quad \mathbb{E}|\tilde{\varepsilon}_N|^2 \leq N^{-1}.$$

For N greater than $(t-s)^{-1}$, Y_N is independent on X_N and (Y_N) is converging to the $\mathcal{N}(0, t-s)$ law. It follows that (X_N, Y_N) is converging in law to a $\mathcal{N}(0, \Gamma_{s,t})$ law, where $\Gamma_{s,t}$ is the diagonal 2×2 matrix $\mathrm{diag}(s, t-s)$. Applying θ_2^{-1}, we deduce (3.30) for $k=2$. □

The following exercise completes the proof of Proposition 3.6.

Exercise 3.17. Let (X_n), (Y_n) be two sequences of random variables on a separable Banach space E such that (X_n) is converging in law to a random variable X.

1. Let (a_n) be a sequence or real numbers converging to $a \in \mathbb{R}$. Assume $\mathbb{E}\|Y_n\|_E^2 \to 0$ when $[n \to +\infty]$. Show that $(a_n X_n + Y_n)$ is converging in law to aX.

2. Assume that (Y_n) is converging in law to a random variable Y and that X_n and Y_n are independent for all n. Show that (X_n, Y_n) is converging in law to (X, Y).

The solution to Exercise 3.17 is here.

3.4.2 Tightness

Our aim now will be to prove the following

Proposition 3.7. *The sequence (μ_{ξ_N}) is tight on $C([0,1];\mathbb{R})$.*

Admit Proposition 3.7 for the moment. By the Prohorov theorem (Theorem 2.13), there is a subsequence of (μ_{ξ_N}) which is converging weakly to a probability measure μ on $C([0,1];R)$. By Proposition 3.6, we have $\mu = \mathbb{P}_W$. This proves the existence of the Wiener measure, on the one-hand, the convergence of the whole sequence (μ_{ξ_N}) by uniqueness of the accumulation points on the other hand. We have therefore these two fundamental corollaries.

Theorem 3.8 (Wiener Measure)**.** *On $C([0,1];\mathbb{R})$ endowed with the Borel σ-algebra, there exists a Wiener measure $\mathbb{P}_W$.*

Theorem 3.9 (Donsker's Theorem)**.** *Let $Z_1, Z_2, \dots$ be independent identically distributed random variables, centred, with variance σ^2. Let $(\xi_N(t))$ be the rescaled random walk defined by (3.29). Then (ξ_N) is converging in law on $C([0,1];\mathbb{R})$ to a one-dimensional Wiener process.*

The extension to the d-dimensional case is straightforward by considering processes with independent coordinates. To prove Proposition 3.7, we will use the Kolmogorov's continuity criterion.

Proof of Proposition 3.7. We do the proof in the case $Z = \pm 1$ with equi-probability. We show first that there exists a constant $C \geq 0$ such that

$$\mathbb{E}|\xi_N(t) - \xi_N(s)|^4 \leq C|t-s|^4, \tag{3.32}$$

for all $s, t \in [0,1]$. We consider first the case $s = t_i$, $t = t_j$ with $i < j$. Then

$$|\xi_N(t) - \xi_N(s)|^4 = \frac{1}{N^2}|Z_{i+1} + \dots + Z_j|^4 = \frac{1}{N^2} \sum_{i<l_1,l_2,l_3,l_4 \leq j} Z_{l_1} Z_{l_2} Z_{l_3} Z_{l_4}. \tag{3.33}$$

If the indices l_1, l_2, l_3, l_4 are not all identical, there may be one index different from any of the other ones, say l_4 for example. In that case, we have, by independence,

$$\mathbb{E}\left[Z_{l_1} Z_{l_2} Z_{l_3} Z_{l_4}\right] = \mathbb{E}\left[Z_{l_1} Z_{l_2} Z_{l_3}\right] \mathbb{E}\left[Z_{l_4}\right] = 0.$$

The indices may also be grouped two by two, for example $l_1 = l_2$, $l_3 = l_4$ with $l_1 \neq l_3$. We obtain then

$$\mathbb{E}\left[Z_{l_1} Z_{l_2} Z_{l_3} Z_{l_4}\right] = \mathbb{E}\left[Z_{l_1}^2\right] \mathbb{E}\left[Z_{l_3}^2\right] = 1.$$

These cross products are

$$Z_{i+1}^2(Z_{i+2}^2 + \dots + Z_j^2), Z_{i+2}^2(Z_{i+3}^2 + \dots + Z_j^2), \dots, Z_{j-1}^2 Z_j^2.$$

There are $\frac{1}{2}(j-i)^2$ such indices in the sum in (3.33). If all the indices coincide (which occurs for $(j-i)$ terms of the sum in (3.33)), then

$$\mathbb{E}\left[Z_{l_1} Z_{l_2} Z_{l_3} Z_{l_4}\right] = \mathbb{E}[Z^4] = 1.$$

Consequently (3.33) gives us

$$\mathbb{E}|\xi_N(t)-\xi_N(s)|^4 = \frac{1}{N^2}[\frac{1}{2}(j-i)^2+(j-i)] \leq \frac{3}{2}\frac{1}{N^2}(j-i)^2 = \frac{3}{2}|t-s|^2.$$

To obtain (3.32) for some general points $s,t\in[0,1]$, let us discuss the size of the increment $|t-s|$ compared to $\frac{1}{N}$. If $|t-s|<\frac{1}{N}$, then either both s and t are in the same intervals $[t_i,t_{i+1}]$, in which case

$$\mathbb{E}|\xi_N(t)-\xi_N(s)|^4 = |t-s|^4N^2\mathbb{E}|Z|^4 \leq |t-s|^2,$$

or $s\in[t_i,t_{i+1}]$ and $t\in[t_{i+1},t_{i+2}]$ for a certain $i\in\{0,\dots,N-1\}$. We have then

$$\begin{aligned}\mathbb{E}|\xi_N(t)-\xi_N(s)|^4 =&\mathbb{E}|(t_{i+1}-s)Z_{i+1}-(t-t_{i+1})Z_{i+2}|^4N^2\\ =&[(t_{i+1}-s)^4+(t-t_{i+1})^4+6(t_{i+1}-s)^2(t-t_{i+1})^2]N^2\\ \leq&(t_{i+1}-s)^2+(t-t_{i+1})^2+6(t_{i+1}-s)(t-t_{i+1})\\ \leq&3[(t_{i+1}-s)^2+(t-t_{i+1})^2+2(t_{i+1}-s)(t-t_{i+1})]\\ =&3(t-s)^2.\end{aligned}$$

Assume $|t-s|\geq\frac{1}{N}$ now. Let i,j be such that $s\in[t_i,t_{i+1}]$, $t\in[t_j,t_{j+1}]$. By (3.31), we have

$$\mathbb{E}|\xi_N(t)-\xi_N(s)|^4 \leq 2^4\left[\mathbb{E}|\xi_N(t_j)-\xi_N(t_{i+1})|^4+\mathbb{E}|\varepsilon_N(t)|^4+\mathbb{E}|\tilde{\varepsilon}_N(s)|^4\right],$$

where

$$\tilde{\varepsilon}_N(s)=(t_{i+1}-s)\sqrt{N}Z_{i+1},\quad \varepsilon_N(t)=(t-t_j)\sqrt{N}Z_{j+1}.$$

It follows that

$$\mathbb{E}|\xi_N(t)-\xi_N(s)|^4 \leq 2^4\left[\frac{3}{2}|t_j-t_{i+1}|^2+\frac{2}{N^2}\right] \leq C|t-s|^2.$$

This concludes the proof of the estimate (3.32). Let us now the Kolmogorov's continuity criterion. Let $\alpha\in(0,\frac{1}{2})$. We obtain that $t\mapsto\xi_N(t)$ is in $C^\alpha([0,1])$, which we already know (it is Lipschitz continuous), but also we have the uniform bound

$$\mathbb{E}\|\xi_N\|_{C^\alpha([0,1];\mathbb{R})}\leq C,$$

by (3.26), where C is independent on N. By the Markov inequality, we deduce that

$$\mathbb{P}(\|\xi_N\|_{C^\alpha([0,1];\mathbb{R})}>R)\leq\frac{C}{R}. \tag{3.34}$$

Let $\varepsilon>0$. Let

$$K_R=\{\xi\in C([0,1];\mathbb{R});\|\xi\|_{C^\alpha([0,1];\mathbb{R})}\leq R\}.$$

By Ascoli's theorem, the set K_R is compact in $C([0,1];\mathbb{R})$. Take $R>C\varepsilon^{-1}$. By (3.34), we have

$$\mathbb{P}(\xi_N\in K_R)\geq 1-\varepsilon$$

for all N. This shows that (ξ_N) is tight in $C([0,1];\mathbb{R})$. □

4 Markov Processes

4.1 Markov process

If E is a Banach space, we denote by $\mathrm{BM}(E)$ denote the vector space of bounded Borel-measurable functions on E. We use the following norm on $\mathrm{BM}(E)$:

$$\|\varphi\|_{\mathrm{BM}(E)} = \sup_{x\in E} |\varphi(x)|.$$

Definition 4.1 (Transition function)**.** Let E be a separable Banach space. A collection $\{Q_t; t \geq 0\}$ of functions on $E \times \mathcal{B}(E)$ is called a time homogeneous *transition function* on E if

1. for all $t \geq 0$, for all $x \in E$, $Q_t(x, \cdot)$ is a Borel probability measure on E,
2. for all $x \in E$, $Q_0(x, \cdot) = \delta_x$, the Dirac mass at x,
3. for all $A \in \mathcal{B}(E)$, $(t, x) \mapsto Q_t(x, A)$ is Borel measurable on $\mathbb{R}_+ \times E$,
4. the following Chapman-Kolmogorov relation is satisfied:

$$Q_{t+s}(x, A) = \int_E Q_s(y, A)Q_t(x, dy) \tag{4.1}$$

for all $0 \leq s, t$, $x \in E$, $A \in \mathcal{B}(E)$.

Definition 4.2 (Markov process)**.** Let E be a separable Banach space and let $\{Q_t; t \geq 0\}$ be a time homogeneous transition function on E. An E-valued process $(X_t)_{t\geq 0}$ is a time-homogeneous *Markov process* associated to $\{Q_t; t \geq 0\}$ if

$$\mathbb{E}\left[\varphi(X_{t+s})|\mathcal{F}_t^X\right] = \int_E \varphi(y)Q_s(X_t, dy), \tag{4.2}$$

for all $\varphi \in \mathrm{BM}(E)$, $0 \leq s, t$, where $\mathcal{F}_t^X = \sigma(\{X_s; 0 \leq s \leq t\})$ (see (4.26) below).

The σ-algebra $\mathcal{F}_s^X$ in Definition 4.2 is the σ-algebra of the past, up to time s (see (4.26)). The σ-algebra $\sigma(X_s)$ is the σ-algebra of the present, relatively to time s. It is clear that (4.2) implies

$$\mathbb{E}\left[\varphi(X_{t+s})|\mathcal{F}_t^X\right] = \mathbb{E}\left[\varphi(X_{t+s})|\sigma(X_t)\right], \tag{4.3}$$

for all $\varphi \in \mathrm{BM}(E)$, $s, t \geq 0$. The identity (4.2) can be rewritten

$$\mathbb{E}\left[\varphi(X_{t+s})|\mathcal{F}_t^X\right] = (P_s\varphi)(X_t), \tag{4.4}$$

where we have introduced the *transition operator* P_t associated to Q_t, defined by

$$P_t\varphi(x) = \int_E \varphi(y)Q_t(x, dy), \quad \varphi \in \mathrm{BM}(E). \tag{4.5}$$

Example 1: process with independent increments. Let $(X_t)_{t\geq 0}$ be a process with independent increments. Let $(\mathcal{F}_t)_{t\geq 0}$ be the natural filtration of the process. For $s,t\geq 0$ and $\varphi\in \mathrm{BM}(E)$, we have

$$\mathbb{E}\left[\varphi(X_{t+s})|\mathcal{F}_t\right] = \mathbb{E}\left[\varphi(X_{t+s}-X_t+X_t)|\mathcal{F}_t\right] = \psi(X_t) \tag{4.6}$$

by independence, where

$$\psi(x) = \mathbb{E}\varphi(X_{t+s}-X_t+x). \tag{4.7}$$

If the increments are i.i.d., we obtain $\psi(x) = \langle Q_s(x,\cdot),\varphi\rangle = P_s\varphi(x)$, where

$$Q_s(x,A) = \mathbb{P}(X_{t+s}-X_t+x\in A). \tag{4.8}$$

Let us check the Chapman-Kolmogorov relation (4.1) under the form $P_t\circ P_s = P_{t+s}$: for $x\in E$, and $0\leq\tau\leq t+s$, we have

$$\begin{aligned}P_{s+t}\varphi(x) = \mathbb{E}\varphi(X_{t+s}-X_0+x) &= \mathbb{E}\left[\mathbb{E}[\varphi(X_{t+s}-X_\tau+X_\tau-X_0+x)|\mathcal{F}_\tau]\right]\\ &= \mathbb{E}\left[(P_{t+s-\tau}\varphi)(X_\tau-X_0+x)\right] = P_\tau\circ P_{t+s-\tau}\varphi(x).\end{aligned}$$

We obtain the result with $\tau=t$. Consequently, $(X_t)_{t\geq 0}$ is a time-homogeneous Markov process. In particular, the Wiener process and the Poisson process are examples of homogeneous Markov processes.

Example 2: the Wiener process. Let $(X_t)_{t\geq 0}$ be a d-dimensional Wiener process. We have

$$P_t\varphi(x) := \int_{\mathbb{R}^d}\varphi(x-y)e^{-\frac{|y|^2}{2t}}\frac{dy}{(2\pi t)^{d/2}} = K_t*\varphi(x), \tag{4.9}$$

where K_t is (up to a coefficient $\frac{1}{2}$) the heat kernel.

Example 3: the Poisson process. Let $(N_t)_{t\geq 0}$ be a Poisson process of exponent $\lambda>0$. By (3.2), we have, for $x\in\mathbb{N}$,

$$P_t\varphi(x) := e^{-\lambda t}\sum_{y\in\mathbb{N}}\varphi(x+y)\frac{[\lambda t]^y}{y!}. \tag{4.10}$$

4.2 Finite-dimensional distributions of a Markov process

Proposition 4.1. *Let E be a separable Banach space, let $(X_t)_{t\geq 0}$ be a time-homogeneous Markov process with transition function $\{Q_t; t\geq 0\}$ and transition operator P_t. Let $\mu_0 = \mathrm{Law}(X_0)$. Then $(P_t)_{t\geq 0}$ and μ_0 determine the finite-dimensional distributions of $(X_t)_{t\geq 0}$.*

Proof of Proposition 4.1. To prove the result, we will establish the following formulae:

$$\mu_{t+s} = (P_t)^*\mu_s, \tag{4.11}$$

and

$$\mu_{t_1,\ldots,t_n} = (P_{t_n-t_{n-1}})^*\otimes\cdots(P_{t_2-t_1})^*\otimes\mu_{t_1}, \tag{4.12}$$

for all $s, t \geq 0$ and $0 \leq t_1 \leq \cdots \leq t_n$. We have introduced the following notations: P_t^* is the dual operator to P_t defined as follows: given μ a Borel probability measure on E (denoted $\mu \in \mathcal{P}(E)$ below), $P_t^* \mu$ is the probability measure defined by

$$P_t^* \mu(A) = \int_E Q_t(x, A) d\mu(x) = \int_E P_t \mathbf{1}_A d\mu.$$

In (4.12), by $\mu_{t_1,\dots,t_n}$, we denote the law of $(X_{t_1}, \dots, X_{t_n})$, an element of $\mathcal{P}(E^n)$. If $\mu_n \in \mathcal{P}(E^n)$ and $\mu_{n+1} \in \mathcal{P}(E^{n+1})$, we say that $\mu_{n+1} = P_t^* \otimes \mu_n$ if

$$\langle \mu_{n+1}, \varphi_1 \otimes \cdots \otimes \varphi_{n+1} \rangle = \langle \mu_n, \varphi_1 \otimes \cdots \otimes (\varphi_n P_t \varphi_{n+1}) \rangle,$$

where

$$(\varphi_1 \otimes \cdots \otimes \varphi_m)(x_1, \dots, x_m) = \varphi_1(x_1) \cdots \varphi_m(x_m).$$

To establish (4.11), note that

$$\langle \mu_t, \varphi \rangle = \mathbb{E}\varphi(X_t) = \mathbb{E}(\mathbb{E}[\varphi(X_t) | \mathcal{F}_s^X]) = \mathbb{E}P_{t-s}\varphi(X_s) = \langle \mu_s, P_{t-s}\varphi \rangle.$$

We establish (4.12) by recursion on n. For $n = 2$, we have

$$\begin{aligned}
\langle \mu_{t_1,t_2}, \varphi_1 \otimes \varphi_2 \rangle =& \mathbb{E}[\varphi_1(X_{t_1})\varphi_2(X_{t_2})] \\
=& \mathbb{E}(\mathbb{E}[\varphi_1(X_{t_1})\varphi_2(X_{t_2}) | \mathcal{F}_{t_1}^X]) \\
=& \mathbb{E}(\varphi_1(X_{t_1}) P_{t_2-t_1}\varphi_2(X_{t_1})) = \langle \mu_{t_1}, \varphi_1 P_{t_2-t_1} \varphi_2 \rangle.
\end{aligned}$$

The proof of $n \mapsto n+1$ in (4.12) is similar. Once (4.12) is established, we express μ_{t_1} in function of μ_0 by (4.11): $\mu_{t_1} = P_{t_1}^* \mu_0$. This concludes the proof. □

As a corollary of Proposition 4.1, we obtain the following result.

Theorem 4.2 (Stationary Markov process)**.** *Let $(X_t)_{t \geq 0}$ be an homogeneous Markov process. Assume that its law is invariant: μ_{X_t} is independent on t. Then $(X_t)_{t \geq 0}$ is stationary.*

Exercise 4.3. Give the proof of Theorem 4.2.

The solution to Exercise 4.3 is here.

4.3 A class of contraction semi-groups

Definition 4.4 (π-convergence)**.** Let E be a separable Banach space. We say that a sequence (φ_n) of $\mathrm{BM}(E)$ is π-converging to $\varphi \in \mathrm{BM}(E)$ (denoted $\varphi_n \xrightarrow{\pi} \varphi$) if $\sup_n \|\varphi_n\|_{\mathrm{BM}(E)} < +\infty$ and $\varphi_n(x) \to \varphi(x)$ for all $x \in E$.

Remark 4.1. This mode of convergence is sometimes called bounded pointwise convergence, b.p.c. (*e.g.* in [EK86, p. 111]).

Let E be a separable Banach space, let $(X_t)_{t\geq 0}$ be a time-homogeneous Markov process with transition function $\{Q_t; t \geq 0\}$ and transition operator P_t. Note that P_t has the following property:

$$\text{if } \varphi_n \xrightarrow{\pi} \varphi, \text{ then } P_t\varphi_n \xrightarrow{\pi} P_t\varphi. \tag{4.13}$$

This is a consequence of the definition (4.5) and the dominated convergence theorem.

Definition 4.5 (π-contraction semi-group, [Pri99])**.** A semi-group of operators $(P_t)_{t\geq 0}$ on $\mathrm{BM}(E)$ is said to be a π-contraction semi-group if $P_0\varphi = \varphi$ and

1. for all $\varphi \in C_b(E)$, for all $x \in E$, $t \mapsto P_t\varphi(x)$ is continuous from the right on $\mathbb{R}_+$,
2. for all $t \geq 0$, P_t has the continuity property (4.13),
3. for all $t \geq 0$, $\|P_t\| \leq 1$ in operator norm.

Note that, in [Pri99], semi-groups $P_t \colon C_b(E) \to C_b(E)$ are considered and $t \mapsto P_t\varphi(x)$ is assumed to be continuous (not just continuous from the right) on $\mathbb{R}_+$. We have modified slightly the notion of π-semi-group introduced in [Pri99], because it gets easier then to compare π-contraction semi-groups and Markov semi-groups. This is the object of the following Proposition 4.3.

Definition 4.6 (Markov semi-group)**.** A semi-group (P_t) of operators on $\mathrm{BM}(E)$ is said to be a *Markov semi-group* if (4.5) is satisfied for a given transition function $\{Q_t\}$.

Proposition 4.3 (π-contraction semi-group and transition semi-groups)**.** *Let $(P_t)_{t\geq 0}$ be a contraction semi-group on $\mathrm{BM}(E)$. We have the following results:*

1. *if $(P_t)_{t\geq 0}$ is a π-contraction semi-group that preserves the positivity ($\varphi \geq 0$ implies $P_t\varphi \geq 0$) and fixes the constants ($P_t\mathbf{1} = \mathbf{1}$), then $(P_t)_{t\geq 0}$ is a Markov semi-group,*
2. *if $(P_t)_{t\geq 0}$ is a Markov semi-group satisfying (4.16) for all $\varphi \in C_b(E)$, then $(P_t)_{t\geq 0}$ is a π-contraction semi-group.*

Remark 4.2 (Stochastic continuity). The property "(4.16) for all $\varphi \in C_b(E)$" is called *stochastic continuity*. If (P_t) is a Markov semi-group, we do not expect (4.16) to be satisfied for all $\varphi \in \mathrm{BM}(E)$. For example if (P_t) is the Heat semi-group,

$$P_t\varphi(x) = \int_{\mathbb{R}^d} \varphi(y)\frac{e^{-\frac{|x-y|^2}{2t}}}{(2\pi t)^{d/2}}dy,$$

then $P_t\varphi(x) \to \varphi(x)$ will only be true for almost every x for general bounded measurable functions. In particular, if $\varphi = \mathbf{1}_A$ where A is a set of measure zero, then $P_t\varphi(x) \to 0$ for every x, thus $P_t\varphi(x)$ does not converge to $\varphi(x)$ when $x \in A$. The Heat semi-group has the property of stochastic continuity however.

Proof of Proposition 4.3. Assume that $(P_t)_{t\geq 0}$ is a π-contraction semi-group that preserves the positivity ($\varphi \geq 0$ implies $P_t\varphi \geq 0$) and fixes the constants. Set $Q_t(x,A) = P_t\mathbf{1}_A(x)$ for $t \in \mathbb{R}_+$, $x \in E$, $A \in \mathcal{B}(E)$. We have several points to consider.
Probability measure. The set function $A \mapsto Q_t(x,A)$ is a probability measure. Indeed $Q_t(x,A) \geq 0$ since $\mathbf{1}_A \geq 0$, $Q_t(x,E) = 1$ since $P_t\mathbf{1} = \mathbf{1}$ and we will see that the property of σ-additivity is satisfied. Let $A_1, A_2, \dots$ be disjoint Borel subsets of E. We have then

$$Q_t(x, A_1 \cup \cdots \cup A_N) = P_t(\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_N})(x) = \sum_{n=1}^{N} Q_t(x, A_n). \tag{4.14}$$

The right-hand side of (4.14) is converging to $\sum_n Q_t(x,A_n)$ when $N \to +\infty$. The left-hand side of (4.14) is $P_t\varphi_N(x)$, where $\varphi_N = \mathbf{1}_{A_1\cup\cdots\cup A_N}$ is π-converging to $\mathbf{1}_A$, $A = \cup_n A_n$. Therefore $P_t\varphi_N \xrightarrow{\pi} P_t\mathbf{1}_A$ by (4.13), and we obtain the countable additivity. Similarly, using the continuity property (4.13), and approaching $\varphi \in \mathrm{BM}(E)$ by a sequence of simple functions, we deduce from the relation $Q_t(\cdot, A) = P_t\mathbf{1}_A$ that (4.5) is satisfied. We have also $Q_0(x,\cdot) = \delta_x$ since $P_0\varphi = \varphi$.
Measurability. Let $A \in \mathcal{B}(E)$. We want to show that $(t,x) \mapsto Q_t(x,A)$ is measurable. The Radon measure $Q_t(x,\cdot)$ is inner regular [Bil99, Theorem 1.1]: $Q_t(x,A) = \sup Q_t(x,F)$, where the supremum is taken over closed sets $F \subset A$. Therefore it is sufficient to consider the case A closed. If A is closed, there is a sequence (φ_k) of Lipschitz bounded functions that π-converges to $\mathbf{1}_A$ (this fact was established in the proof of Proposition 2.8, see also Remark 2.9). Consequently $Q_t(x,A)$ is the limit of $P_t\varphi_k(x)$ when $k \to +\infty$ and that $(t,x) \mapsto Q_t(x,A)$ is measurable follows from the fact that $(t,x) \mapsto P_t\varphi(x)$ is measurable when $\varphi \in C_b(E)$. Indeed, the map $h\colon (t,x) \mapsto P_t\varphi(x)$ is continuous from the right in t and measurable in x. Consider a regular partition of $\mathbb{R}_+ \setminus \{0\}$ in intervals $(a,b]$ of length N^{-1} and approximate $h(\cdot,x)$ on $(a,b]$ by the value $h(b,x)$ at the right of the interval, we obtain[4] a sequence of $\mathcal{B}(\mathbb{R}_+ \times E)$- measurable functions h_N that π-converges to h.
Chapman-Kolmogorov property. The Chapman-Kolmogorov property (4.1) follows from the semi-group property of $(P_t)_{t\geq 0}$ and (4.5).
Conversely, assume now that $(P_t)_{t\geq 0}$ is a semi-group of transition operators with the continuity property (4.16) for all $\varphi \in C_b(E)$. We have seen that $(P_t)_{t\geq 0}$ satisfies (4.16). By (4.13) and the semi-group property, (4.16) implies condition 1 in Definition 4.5. This proves the result. □

Remark 4.3 (Feller-semi-group). A contraction semi-group $(P_t)_{t\geq 0}$ on $\mathrm{BM}(E)$ is said to be Feller if $C_b(E)$ is stable by P_t. A simple example of non-Feller contraction semi-group is given by

$$P_t\varphi = e^{-t}\varphi + (1 - e^{-t})\langle \varphi, \nu\rangle \psi,$$

where ν is a probability measure on E and ψ a function in $\mathrm{BM}(E) \setminus C_b(E)$ such that $\langle \psi, \nu\rangle = 1$.

[4] we also set $h_N(0,x) = h(0,x)$

4.4 Infinitesimal generator

Let (P_t) be a π-contraction semi-group. We define the *infinitesimal generator* $\mathcal{L}$ of (P_t) as follows: $\varphi \in C_b(E)$ is in the *domain* $\mathcal{D}(\mathcal{L})$ of $\mathcal{L}$ if there exists $\psi \in \mathrm{BM}(E)$ such that

$$\frac{P_t\varphi - \varphi}{t} \xrightarrow{\pi} \psi. \tag{4.15}$$

We then set $\mathcal{L}\varphi = \psi$. Note that if $\varphi \in \mathcal{D}(\mathcal{L})$, then

$$P_t\varphi \xrightarrow{\pi} \varphi \tag{4.16}$$

when $t \to 0$.

Proposition 4.4. *For all $t \geq 0$, for all $\varphi \in \mathcal{D}(\mathcal{L})$, we have $P_t\varphi \in \mathcal{D}(\mathcal{L})$ and $\mathcal{L}P_t\varphi = P_t\mathcal{L}\varphi$. Besides, for all $x \in E$, the map $t \mapsto P_t\varphi(x)$ from $\mathbb{R}_+$ to $\mathbb{R}$ is differentiable on $\mathbb{R}_+$, with*

$$\frac{d}{dt}P_t\varphi(x) = \mathcal{L}P_t\varphi(x) = P_t\mathcal{L}\varphi(x). \tag{4.17}$$

Proof of Proposition 4.4. It results from the semi-group property $P_{t+s} = P_t \circ P_s$, which gives

$$\frac{P_sP_t\varphi - P_t\varphi}{s} = \frac{P_{s+t}\varphi - P_t\varphi}{s} = P_t\frac{P_s\varphi - \varphi}{s}, \tag{4.18}$$

and from the continuity property (4.13). □

Remark 4.4 (Strongly continuous semi-groups). If

$$\lim_{t\to 0} \|P_t\varphi - \varphi\|_{\mathrm{BM}(E)} = 0 \tag{4.19}$$

for all $\varphi \in C_b(E)$, then $(P_t)_{t\geq 0}$ is a C_0 semi-group on G_E [Paz83, p. 4]. *This will generally not be the case unless E has finite dimension.* We can then define the infinitesimal generator $\mathcal{L}$ by considering the limit of $\frac{P_t\varphi-\varphi}{t}$ in $\mathrm{BM}(E)$ (for the sup norm hence). The Hille-Yosida theorem [Paz83, p. 8] characterizes the unbounded operators $\mathcal{L}$ which give rise to a C_0-semi-group of contraction.

Lemma 4.5. *Let $\theta\colon \mathbb{R}_+ \to \mathbb{R}_+$ be an integrable function of class C^1 such that θ' is integrable. Suppose that the semi-group of transition operators $(P_t)_{t\geq 0}$ satisfies (4.16) for all $\varphi \in C_b(E)$. Then*

$$\psi_\theta := \int_0^\infty \theta(t)P_t\varphi dt \in \mathcal{D}(\mathcal{L}), \quad \mathcal{L}\psi_\theta = -\theta(0)\varphi - \int_0^\infty \theta'(t)P_t\varphi dt, \tag{4.20}$$

for all $\varphi \in C_b(E)$.

Proof of Lemma 4.5. The function ψ_θ is well defined in $\mathrm{BM}(E)$: it is measurable as the sum of measurable quantities, and bounded since $|\psi_\theta(x)| \leq \|\varphi\|_{\mathrm{BM}(E)}\|\theta\|_{L^1(\mathbb{R}_+)}$. We compute

$$\begin{aligned}\frac{1}{s}(P_s\psi_\theta - \psi_\theta) =& \frac{1}{s}\left[\int_0^\infty \theta(t)P_{t+s}\varphi dt - \int_0^\infty \theta(t)P_t\varphi dt\right] && (4.21)\\ =& \frac{1}{s}\left[\int_s^\infty \theta(t-s)P_t\varphi dt - \int_0^\infty \theta(t)P_t\varphi dt\right]\\ =& \int_s^\infty \frac{\theta(t-s)-\theta(t)}{s}P_t\varphi dt - \frac{1}{s}\int_0^s \theta(t)P_t\varphi dt. && (4.22)\end{aligned}$$

The first term in (4.22) π-converges to $-\int_0^\infty \theta'(t)P_t\varphi dt$, the second term π-converges to $-\theta(0)\varphi$ by (4.16). To obtain (4.21), we have cut the integral at level n and used the estimates

$$P_s\int_n^\infty \theta(t)P_t\varphi dt, \quad \int_n^\infty \theta(t)P_{t+s}\varphi dt = \mathcal{O}(\|\theta\|_{L^1(n,+\infty)}\|\varphi\|_{\mathrm{BM}(E)})$$

in the $\mathrm{BM}(E)$-norm to neglect the remainder terms at the limit $n \to +\infty$. □

As a corollary to (the proof) of Lemma 4.5, we have the following result.

Proposition 4.6. *Suppose that the semi-group of transition operators $(P_t)_{t\geq 0}$ satisfies (4.16). Then the domain $\mathcal{D}(\mathcal{L})$ is π-dense in $C_b(E)$.*

Proof of Proposition 4.6. if $\varphi \in C_b(E)$, we have

$$\frac{1}{t}\int_0^t P_s\varphi ds \in \mathcal{D}(\mathcal{L}), \quad \frac{1}{t}\int_0^t P_s\varphi ds \xrightarrow{\pi} \varphi \tag{4.23}$$

when $t \to 0$. □

We will apply Lemma 4.5 with $\theta(t) = e^{-\lambda t}$, $\lambda > 0$ in particular. We denote then by

$$R_\lambda\varphi = \int_0^\infty e^{-\lambda t}P_t\varphi dt, \tag{4.24}$$

the *resolvent* of (P_t). Here, (4.20) gives the identity $\mathcal{L}R_\lambda\varphi = \lambda R_\lambda\varphi - \varphi$, *i.e.* for $\varphi \in C_b(E)$,

$$R_\lambda\varphi \in \mathcal{D}(\mathcal{L}), \quad (\lambda - \mathcal{L})R_\lambda\varphi = \varphi. \tag{4.25}$$

4.5 Filtration

Definition 4.7 (Filtration)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A family $(\mathcal{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathcal{F}$ is said to be a *filtration* if the family is increasing with respect to t: $\mathcal{F}_s \subset \mathcal{F}_t$ for all $0 \leq s \leq t$. The space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a filtered space. If $(\mathcal{F}_t)_{t\geq 0}$ we set $\mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$. We say that $(\mathcal{F}_t)_{t\geq 0}$ is *continuous from the right* if $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t. We say that $(\mathcal{F}_t)_{t\geq 0}$ is *complete* if $\mathcal{F}_t$ is complete: it contains all $\mathbb{P}$-negligible sets. We say that $(\mathcal{F}_t)_{t\geq 0}$ *satisfies the usual condition* if $(\mathcal{F}_t)_{t\geq 0}$ is continuous from the right and complete.

Definition 4.8 (Adapted process)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and E a separable Banach space. An E-valued process $(X_t)_{t\geq 0}$ is said to be *adapted* if, for all $t \geq 0$, X_t is $\mathcal{F}_t$-measurable.

Note that this means $\sigma(X_t) \subset \mathcal{F}_t$ for all $t \geq 0$.

Example 4.5. If $(X_t)_{t\geq 0}$ is a process over $(\Omega, \mathcal{F}, \mathbb{P})$, we introduce

$$\mathcal{F}_t^X = \sigma(\{X_s; 0 \leq s \leq t\}) \tag{4.26}$$

the σ-algebra generated by all random variables $(X_{s_1}, \ldots, X_{s_N})$ for $N \in \mathbb{N}^*$, $s_1, \ldots, s_N \in [0, t]$. Then $(\mathcal{F}_t^X)_{t\geq 0}$ is a filtration and $(X_t)_{t\geq 0}$ is adapted to this filtration: $(\mathcal{F}_t^X)_{t\geq 0}$ is called the natural filtration of the process, or the filtration generated by $(X_t)_{t\geq 0}$.

Exercise 4.9. Let $(X_t)_{t\geq 0}$ be a continuous process adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Show that $(\mathcal{F}_t^X)_{t\geq 0}$ is not necessarily continuous from the right. *Hint:* you may consider $X_t = tY$, Y being given. *The solution to Exercise 4.9 is here.*

Proposition 4.7. *We assume that $(\mathcal{F}_t)$ is complete. Then any limit (a.s., or in probability, or in $L^p(\Omega)$) of adapted processes is adapted.*

Proof of Proposition 4.7. Let X^n and X be some E-valued random variables such that $(X^n)_{n\in\mathbb{N}}$ is converging to X for one of the modes of convergence that we are considering. We just have to consider convergence almost-sure since convergence in probability or in $L^p(\Omega)$ implies convergence a.s. of a subsequence. If all the X^n are $\mathcal{G}$-measurable, where $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then the set of points where (X_n) is converging is in $\mathcal{G}$ (we use the Cauchy criterion to characterize the convergence). Consequently, X is equal $\mathbb{P}$-a.e. to a $\mathcal{G}$-measurable function. If $\mathcal{G}$ is complete, we deduce that X is $\mathcal{G}$-measurable. □

Definition 4.10 (Markov process relatively to a filtration)**.** Let E be a separable Banach space. Let $\{Q_t; t \geq 0\}$ be a transition function on E and let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. An E-valued process $(X_t)_{t\geq 0}$ is a time-homogeneous *Markov process relatively to* $(\mathcal{F}_t)_{t\geq 0}$ associated to $\{Q_t; t \geq 0\}$ if

$$\mathbb{E}\left[\varphi(X_{t+s})|\mathcal{F}_t\right] = \int_E \varphi(y) Q_s(X_t, dy), \tag{4.27}$$

for all $\varphi \in \mathrm{BM}(E)$, $0 \leq s, t$.

The Markov property (4.27) with respect to $(\mathcal{F}_t)_{t\geq 0}$ implies (4.2) with respect to the filtration $(\mathcal{F}_t^X)_{t\geq 0}$. Indeed, (4.27) implies that X_t is $\mathcal{F}_t$-measurable, hence $\mathcal{F}_t^X \subset \mathcal{F}_t$. We can then deduce (4.2) from the identity

$$\mathbb{E}\left[\varphi(X_{t+s})|\mathcal{F}_t^X\right] = \mathbb{E}\left[\mathbb{E}\left[\varphi(X_{t+s})|\mathcal{F}_t\right]|\mathcal{F}_t^X\right]$$

and from (4.27) thus. However, it is necessary to extend Definition 4.2 into Definition 4.10, for at least two reasons: 1. it is sometimes easier to prove (4.27) (see Exercise 4.11 below for example), 2. the strong Markov property (see Section 4.6) naturally involves stopping times with respect to certain filtrations which have no reasons to be the filtration $(\mathcal{F}_t^X)_{t\geq 0}$.

Exercise 4.11 (Markov jump process)**.** Let $(X_n)_{n\geq 0}$ be a discrete time-homogeneous Markov chain on E with transition function Q_n and transition operator P_n, $n \in \mathbb{N}$. Let $N(t)$ be a Poisson process of exponent 1 independent on $(X_n)_{n\geq 0}$ and let $\xi_t = X_{N(t)}$. Introduce also $\mathcal{F}_t = \mathcal{F}_t^\xi \vee \mathcal{F}_t^N$, the minimal σ-algebra containing $\mathcal{F}_t^\xi$ and $\mathcal{F}_t^N$.

1. Show that $P_n = P_1^n$ for all $n \geq 0$.
2. Show that, for all $E \in \mathcal{F}_t$,
$$\mathbb{E}\left[\mathbf{1}_E \varphi(X_{n+N(t)})\right] = \mathbb{E}\left[\mathbf{1}_E P_n \varphi(X_{N(t)})\right]. \tag{4.28}$$
Hint: you may prove (4.28) for E of the form $E = B \cap D \cap \{N(t) = m\}$, $m \in \mathbb{N}$, $B \in \mathcal{F}_m^X$, $D \in \mathcal{F}_t^N$ first.
3. Show that $(\xi_t)_{t\geq 0}$ is a time-homogeneous Markov process with respect to $(\mathcal{F}_t)_{t\geq 0}$ with transition function
$$\rho_t(x, A) = e^{-t} \sum_{n\geq 0} \frac{t^n}{n!} Q_n(x, A) \tag{4.29}$$
and transition operator and infinitesimal generator
$$\Pi_t = e^{-t(\mathrm{Id}-P_1)}, \quad \mathcal{L} = P_1 - \mathrm{Id}. \tag{4.30}$$

The solution to Exercise 4.11 is here.

4.6 Stopping time and strong Markov property

Definition 4.12 (Stopping time)**.** Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. A random variable τ with values in $[0, +\infty]$ is an $(\mathcal{F}_t)$-stopping time (or *stopping time* relatively to $(\mathcal{F}_t)_{t\geq 0}$) if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. If τ is a stopping time, we denote by $\mathcal{F}_\tau$ the σ-algebra
$$\mathcal{F}_\tau = \{A \in \mathcal{F}; A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}. \tag{4.31}$$

Remark 4.6. If $\mathcal{F}_t$ describes the information accessible at time t, $\mathcal{F}_\tau$ describes the information accessible (via the filtration) by the knowledge of τ. This last statement is quite informal. To give a more rigorous version of it, consider the case of a *discrete* stopping time τ: we assume that τ takes its values in the finite set $\{t_1, \ldots, t_m\}$, with $0 \leq t_1 < \cdots < t_m$. To decide if an event A is in $\mathcal{F}_\tau$, we look at $A \cap \{\tau = t_i\}$: this should be in $\mathcal{F}_{t_i}$. Observe that requiring $A \cap \{\tau = t_i\} \in \mathcal{F}_{t_i}$ for all i is equivalent to the requirement in (4.31).

Exercise 4.13. Let E be a separable Banach space. Let (X_t) be an E-valued stochastic process. Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. Recall that $\mathcal{F}_{t+} = \cap_{s>t} \mathcal{F}_s$.

1. Show that τ is a stopping time relatively to $(\mathcal{F}_{t+})_{t\geq 0}$ if, and only if, $\{\tau < t\} \in \mathcal{F}_t$ for all $t > 0$.

2. Show that, for all $s \geq 0$, $\tau \wedge s$ is a stopping time, that $\tau \wedge s$ is $\mathcal{F}_s$-measurable and that $\mathcal{F}_{\tau \wedge s} \subset \mathcal{F}_s$.

3. We assume (X_t) continuous and adapted to $(\mathcal{F}_t)_{t\geq 0}$. Let A be a *closed* set. Show that the hitting time
$$\tau_A = \inf\{t \geq 0; X_t \in A\} \tag{4.32}$$
is an $(\mathcal{F}_t)$-stopping time. *Hint:* consider $t \mapsto d(X_t, A) = \inf_{y\in A} \|X_t - y\|_E$.

4. We assume that (X_t) is continuous from the right and adapted to $(\mathcal{F}_t)_{t\geq 0}$. Let A be an *open* set. Show that the hitting time
$$\tau_A = \inf\{t \geq 0; X_t \in A\} \tag{4.33}$$
is an $(\mathcal{F}_{t+})$-stopping time.

5. Let τ be a discrete $(\mathcal{F}_t^X)$-stopping time. Show that
$$\mathcal{F}_\tau = \sigma(\{X(t \wedge \tau); t \geq 0\}). \tag{4.34}$$

The solution to Exercise 4.13 is here.

Definition 4.14 (Progressively measurable process)**.** Let $(\mathcal{F}_t)_{t\in[0,T]}$ be a filtration. An E-valued process $(X_t)_{t\in[0,T]}$ is said to be *progressively measurable* (with respect to $(\mathcal{F}_t)_{t\in[0,T]}$) if, for all $t \in [0, T]$, the map $(s, \omega) \mapsto X_s(\omega)$ from $[0, t] \times \Omega$ to E is $\mathcal{B}([0, t]) \times \mathcal{F}_t$-measurable.

Definition 4.15 (Strong Markov property)**.** Let E be a separable Banach space, let $(X_t)_{t\geq 0}$ be a time-homogeneous Markov process with transition function $\{Q_t; t \geq 0\}$ and transition operator P_t. We assume that (X_t) is progressively measurable. Let τ be a stopping time such that $\tau < +\infty$ a.s. We say that (X_t) is *strong Markov* at τ if
$$\mathbb{E}[\varphi(X_{\tau+t})|\mathcal{F}_\tau] = (P_t\varphi)(X_\tau), \tag{4.35}$$
for all $\varphi \in \mathrm{BM}(E)$.

Note that X_τ is a random variable that is $\mathcal{F}_\tau$-measurable, due to the fact that (X_t) is progressively measurable. Indeed, if $t \geq 0$, Question 2 in Exercise 4.13 shows that $\tau \wedge t$ is $\mathcal{F}_t$-measurable. The map $\omega \mapsto X(\tau(\omega) \wedge t, \omega)$ is the composition of $(s, \omega) \mapsto X(s, \omega)$ from $[0, t] \times \Omega$, that is $\mathcal{B}([0, t]) \times \mathcal{F}_t$-measurable, with the measurable map $\omega \mapsto (\tau(\omega) \wedge t, \omega)$ from Ω endowed with $\mathcal{F}_t$) to $[0, t] \times \Omega$ endowed with $\mathcal{B}([0, t]) \times \mathcal{F}_t$. Consequently, $X_{\tau\wedge t}$ is $\mathcal{F}_t$-measurable. If $B \in \mathcal{B}(E)$ it follows that
$$\{X_\tau \in B\} \cap \{\tau \leq t\} = \{X_{\tau\wedge t} \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t.$$
This proves the result.

Proposition 4.8 (Strong Markov property - discrete time)**.** *Let E be a separable Banach space, let $(X_t)_{t\geq 0}$ be a time-homogeneous Markov process with transition function $\{Q_t; t \geq 0\}$ and transition operator P_t. We assume that (X_t) is progressively measurable. Let τ be a* discrete *stopping time such that $\tau < +\infty$ a.s. Then (X_t) is strong Markov at τ.*

Proof of Proposition 4.8. Let $0 \leq t_1 < \cdots < t_m$ be the values taken by τ. Let $\varphi \in \mathrm{BM}(E)$ and let $B \in \mathcal{F}_\tau$. We have

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{B\cap\{\tau=t_i\}}\varphi(X_{\tau+t})\right] &= \mathbb{E}\left[\mathbf{1}_{B\cap\{\tau=t_i\}}\varphi(X_{t_i+t})\right] \\ &= \mathbb{E}\left[\mathbf{1}_{B\cap\{\tau=t_i\}}\mathbb{E}[\varphi(X_{t_i+t})|\mathcal{F}_{t_i}]\right] && (4.36) \\ &= \mathbb{E}\left[\mathbf{1}_{B\cap\{\tau=t_i\}}(P_t\varphi)(X_{t_i})\right] && (4.37) \\ &= \mathbb{E}\left[\mathbf{1}_{B\cap\{\tau=t_i\}}(P_t\varphi)(X_{\tau})\right]. && (4.38)\end{aligned}$$

The identity (4.36) is due to the fact that $B \cap \{\tau = t_i\} \in \mathcal{F}_{t_i}$, (4.37) uses the (standard) Markov property. Summing (4.38) over i, we get the result. □

5 Martingale

Definition 5.1 (Martingale)**.** Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered space and E a separable Banach space. Let $(X_t)_{t\geq 0}$ be a L^1, E-valued process: for all $t \geq 0$, $X_t \in L^1(\Omega)$. The process $(X_t)_{t\geq 0}$ is said to be a *martingale* if, for all $0 \leq s \leq t$, $X_s = \mathbb{E}(X_t|\mathcal{F}_s)$.

Remark 5.1.

1. A martingale with continuous (*resp.*, càdlàg) trajectories is said to be a continuous (*resp.*, càdlàg) martingale.
2. If $(X_t)_{t\geq 0}$ is a martingale, then it is adapted to $(\mathcal{F}_t)_{t\geq 0}$.
3. With respect to a fixed time $t > 0$, conditioning on $\mathcal{F}_s$ with $s \leq t$ is a way to average over all events which occurred between times s and t. For a martingale, this will let the position X_s unchanged. In the scalar case $E = \mathbb{R}$, a process $(X_t)_{t\geq 0}$ is said to be a *sub-martingale* if (X_t) is adapted and X_s is below the average $\mathbb{E}(X_t|\mathcal{F}_s)$ for all $0 \leq s \leq t$. If $X_s \geq \mathbb{E}(X_t|\mathcal{F}_s)$ for all $0 \leq s \leq t$, then an adapted process $(X_t)_{t\geq 0}$ is said to be a *super-martingale*.

Exercise 5.2. Let $(X_t)_{t\geq 0}$ be a real-valued process adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$ such that $X_t - X_s$ is independent on $\mathcal{F}_s$ for all $0 \leq s \leq t$. We assume that $(X_t)_{t\geq 0}$ has finite second moment and is centred: $\mathbb{E}|X_t|^2 < +\infty$, $\mathbb{E}[X_t] = 0$ for all $t \geq 0$.

1. Show that $(X_t)_{t\geq 0}$ is a martingale .
2. Show that $t \mapsto \mathbb{E}[X_t^2]$ is increasing.
3. Show that $(X_t^2 - \mathbb{E}[X_t^2])_{t\geq 0}$ is a martingale.

The solution to Exercise 5.2 is here.

Exercise 5.3. 1. If $\varphi\colon \mathbb{R} \to (-\infty, +\infty]$ is *proper*, *i.e.* $\varphi(x)$ is finite for at least an $x \in \mathbb{R}$, we denote by φ^* the Legendre-Fenchel conjugate of φ defined by

$$\varphi^*(p) = \sup_{x\in\mathbb{R}}[xp - \varphi(x)] \in (-\infty, +\infty].$$

(a) Show that φ^* is convex and continuous.

(b) We admit that, if φ is convex, then $\varphi = \varphi^{**}$. Show that

$$\varphi(x) = \sup_{p\in D}[px - \varphi^*(p)],$$

where D is a countable subset of $\mathbb{R}$.

(c) Let X be a real-valued L^1 random variable, $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$ and $\varphi\colon \mathbb{R} \to \mathbb{R}$ a convex function such that $\varphi(X) \in L^1(\Omega)$. Show that

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}] \text{ a.s.} \tag{5.1}$$

(d) Let $(X_t)_{t\geq 0}$ be a real-valued martingale relatively to a filtration $(\mathcal{F}_t)_{t\geq 0}$. Let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be a convex function such that $\varphi(X_t) \in L^1(\Omega)$ for all $t \geq 0$. Show that $(\varphi(X_t))_{t\geq 0}$ is a sub-martingale.

2. Let E be a Banach space such that the dual E^* is separable. Let X be an E-valued L^1 random variable, $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Show that

$$\|\mathbb{E}[X|\mathcal{G}]\|_E \leq \mathbb{E}[\|X\|_E|\mathcal{G}] \text{ a.s.} \tag{5.2}$$

3. Let E be a Banach space such that the dual E^* is separable. Let $(X_t)_{t\geq 0}$ be an E-valued martingale relatively to a filtration $(\mathcal{F}_t)_{t\geq 0}$. Show that $(\|X_t\|_E)_{t\geq 0}$ is a sub-martingale.

The solution to Exercise 5.3 is here.

Remark 5.2. By Question 1d of Exercise (5.3), if $(X_t)_{t\in[0,T]}$ is a real-valued martingale, then $(X_t^2)_{t\geq 0}$, or, more generally, $(|X_t|^p)_{t\geq 0}$ for $p \geq 1$, is a submartingale.

Lemma 5.1. *Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)$-submartingale. Let τ_1 and τ_2 be two discrete stopping times relatively to $(\mathcal{F}_t)_{t\geq 0}$. Then $X_{\tau_1\wedge\tau_2} \leq \mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}]$.*

Proof of Lemma 5.1. Let $0 \leq t_1 < \cdots < t_m$ be the values taken by τ_1. Given $A \in \mathcal{F}_{\tau_1}$, we want to show that

$$\mathbb{E}\left[\mathbf{1}_A X_{\tau_2}\right] \geq \mathbb{E}\left[\mathbf{1}_A X_{\tau_2\wedge\tau_1}\right]. \tag{5.3}$$

By decomposing $A = \cup_{i=1}^m A \cap \{\tau_1 = t_i\}$, (5.3) is equivalent to

$$\mathbb{E}\left[\mathbf{1}_{A\cap\{\tau_1=t_i\}} X_{\tau_2}\right] \geq \mathbb{E}\left[\mathbf{1}_{A\cap\{\tau_1=t_i\}} X_{\tau_2\wedge\tau_1}\right] = \mathbb{E}\left[\mathbf{1}_{A\cap\{\tau_1=t_i\}} X_{\tau_2\wedge t_i}\right], \tag{5.4}$$

for all $i \in \{1, \ldots, m\}$. Since $A \cap \{\tau_1 = t_i\} \in \mathcal{F}_{t_i}$, (5.4) follows from the inequality

$$\mathbb{E}\left[X_{\tau_2}|\mathcal{F}_{t_i}\right] \geq X_{\tau_2\wedge t_i}. \tag{5.5}$$

To obtain (5.5), we split $\mathbb{E}\left[X_{\tau_2}|\mathcal{F}_{t_i}\right]$ into the sum of the two terms $\mathbb{E}\left[X_{\tau_2}\mathbf{1}_{\tau_2>t_i}|\mathcal{F}_{t_i}\right]$ and $\mathbb{E}\left[X_{\tau_2}\mathbf{1}_{\tau_2\leq t_i}|\mathcal{F}_{t_i}\right]$. The second term is

$$\mathbb{E}\left[X_{\tau_2}\mathbf{1}_{\tau_2\leq t_i}|\mathcal{F}_{t_i}\right] = \mathbb{E}\left[X_{\tau_2\wedge t_i}\mathbf{1}_{\tau_2\leq t_i}|\mathcal{F}_{t_i}\right] = X_{\tau_2\wedge t_i}\mathbf{1}_{\tau_2\leq t_i},$$

since $X_{\tau_2\wedge t_i}$ is $\mathcal{F}_{t_i}$-measurable (Question 2 of Exercise 4.13). For the first term, we have

$$\mathbb{E}\left[X_{\tau_2}\mathbf{1}_{\tau_2>t_i}|\mathcal{F}_{t_i}\right] = \mathbb{E}\left[X_{\tau_2}|\mathcal{F}_{t_i}\right]\mathbf{1}_{\tau_2>t_i}.$$

The conclusion will therefore follow from the inequality

$$\mathbb{E}\left[X_{\tau_2}|\mathcal{F}_{t_i}\right] \geq X_{\tau_2\wedge t_i}, \tag{5.6}$$

Note that (5.6) corresponds to the general inequality $X_{\tau_1\wedge\tau_2} \leq \mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}]$ that we want to prove, in the special case $\tau_1 = t_i$ a.s. To establish (5.6), we denote by $s_1 < \cdots < s_q$ the values taken by τ_2. For $t \in [s_j, s_{j+1}]$, we have

$$\begin{aligned}
\mathbb{E}\left[X_{\tau_2\wedge s_{j+1}}|\mathcal{F}_t\right] &= \mathbb{E}\left[X_{\tau_2\wedge s_{j+1}}(\mathbf{1}_{\tau_2>t} + \mathbf{1}_{\tau_2\leq t})|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[X_{s_{j+1}}|\mathcal{F}_t\right]\mathbf{1}_{\tau_2>t} + X_{s_j}\mathbf{1}_{\tau_2\leq t} \\
&\geq X_t\mathbf{1}_{\tau_2>t} + X_{s_j}\mathbf{1}_{\tau_2\leq t} = X_{\tau_2\wedge t}.
\end{aligned} \tag{5.7}$$

In (5.7), we have used the fact that $(X)_t$ is a sub-martingale. We apply (5.7) with $j = q-1$, $q-2, \ldots$, $t = s_{q-1}$, $t = s_{q-2}, \ldots$. This gives (since $\tau_2 \wedge s_q = \tau_2$)

$$\mathbb{E}\left[X_{\tau_2}|\mathcal{F}_{s_{q-1}}\right] \geq X_{\tau_2\wedge s_{q-1}}, \quad \mathbb{E}\left[X_{\tau_2\wedge s_{q-1}}|\mathcal{F}_{s_{q-2}}\right] \geq X_{\tau_2\wedge s_{q-2}}, \quad \ldots .$$

Using (2.44), we obtain $\mathbb{E}\left[X_{\tau_2}|\mathcal{F}_{s_j}\right] \geq X_{\tau_2\wedge s_j}$ where $j \in \{1,\ldots,q\}$ is such that $s_{j-1} \leq t_i < s_j$. We apply then (5.7) once more with $t = t_i$ and use (2.44) again to obtain (5.6). □

Let us consider the case of general (non necessarily discrete) stopping times. If (X_t) is a right-continuous submartingale, τ_1 and τ_2 are two $(\mathcal{F}_t)$-stopping times, then

$$X(\tau_1\wedge\tau_2\wedge T) \leq \mathbb{E}\left[X(\tau_2\wedge T)|\mathcal{F}_{\tau_1}\right].$$

If in addition, τ_2 is finite a.s., $\mathbb{E}|X(\tau_2)| < +\infty$ and $\lim_{T\to+\infty}\mathbb{E}\left[|X(T)|\mathbf{1}_{T>\tau_2}\right] = 0$, then

$$X(\tau_1\wedge\tau_2) \leq \mathbb{E}\left[X(\tau_2)|\mathcal{F}_{\tau_1}\right]. \tag{5.8}$$

See [EK86, Theorem 2.13 p.61]. If (X_t) is a right-continuous martingale, and all the necessary hypotheses are fulfilled, we can apply (5.8) to $-X$, we obtain thus the equality

$$X(\tau_1\wedge\tau_2) = \mathbb{E}\left[X(\tau_2)|\mathcal{F}_{\tau_1}\right], \tag{5.9}$$

which is the content of the Doob's *optional sampling theorem*.

We will use Lemma 5.1 in the proof of Theorem 5.5. For the moment, we will need the following corollary to Lemma 5.1.

Corollary 5.2. *Let $(X_t)_{t\geq 0}$ be an $(\mathcal{F}_t)$-submartingale taking non-negative values. Let $T > 0$ and let J be a finite subset of $[0,T]$. Then*

$$\mathbb{P}\left(\sup_{t\in J} X_t \geq \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}\left[\mathbf{1}_{\{\sup_{t\in J} X_t\geq\lambda\}} X_T\right]. \tag{5.10}$$

Proof of Lemma 5.2. Define the stopping time $\tau = \min\{t \in J, X_t \geq \lambda\}$, with the usual convention $\tau = +\infty$ if $X_t < \lambda$ for all $t \in J$. We want to prove

$$\mathbb{P}(\mathbf{1}_{\tau<+\infty}) \leq \frac{1}{\lambda}\mathbb{E}\left[\mathbf{1}_{\tau<+\infty} X_T\right]. \tag{5.11}$$

Let $\tau_1 = \tau \wedge T$, $\tau_2 = T$. Note that $\{\tau < +\infty\} \in \mathcal{F}_{\tau_1}$. By Lemma 5.1, we have therefore

$$X_{\tau_1\wedge\tau_2}\mathbf{1}_{\tau<+\infty} \leq \mathbb{E}[X_T\mathbf{1}_{\tau<+\infty}|\mathcal{F}_{\tau_1}]. \tag{5.12}$$

Since $\lambda\mathbf{1}_{\tau<+\infty} \leq X_{\tau_1} = X_{\tau_1\wedge\tau_2}$, taking expectation in (5.12) gives us (5.11). $\square$

Using Corollary 5.2, we will establish the following result.

Theorem 5.3 (Doob's martingale inequality)**.** *Let $p > 1$. Let $(M_t)_{t\in[0,T]}$ be a continuous, real-valued martingale, such that $\mathbb{E}|M_T|^p < +\infty$. Then the inequality*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |M_t|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_T|^p \tag{5.13}$$

is satisfied.

Proof of Theorem 5.3. We admit first the result for discrete-time martingales and simply explain the end of the proof... which is straightforward then. Indeed, by continuity of the process, we have

$$\sup_{t\in[0,T]} |M_t|^p = \sup_{n\geq 1}\sup_{i=1,\dots,n} |M_{t_i}|^p,$$

where $\{t_2,\dots\}$ is an enumeration of $[0,T)\cap\mathbb{Q}$ and $t_1 = T$. This shows first that $\sup_{t\in[0,T]}|M_t|^p$ is measurable, and also gives the result since

$$\mathbb{E}\left[\sup_{i=1,\dots,n} |M_{t_i}|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_T|^p$$

for all n by the discrete-time case. There remains to show the discrete-time case: if $J \subset [0,T]$ is finite, we want to prove that

$$\mathbb{E}\left[(M_J^*)^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_T|^p, \quad M_J^* := \sup_{t\in J}|M_t|. \tag{5.14}$$

We use the fact that $X_t := |M_t|$ is a non-negative sub-martingale. By Corollary 5.2, we have

$$\mathbb{P}(M_J^* > \lambda) \leq \frac{1}{\lambda}\mathbb{E}\left[\mathbf{1}_{M_J^*>\lambda}|M_T|\right].$$

Let $k > 0$. By Fubini's Theorem, we obtain

$$\mathbb{E}[(M_J^* \wedge k)^p] = \mathbb{E}\int_0^k p\lambda^{p-1}\mathbf{1}_{M_J^*>\lambda}d\lambda = \int_0^k p\lambda^{p-1}\mathbb{P}(M_J^* > \lambda)d\lambda, \tag{5.15}$$

and thus

$$\mathbb{E}[(M_J^* \wedge k)^p] \leq \frac{p}{p-1}\mathbb{E}[(M_J^* \wedge k)^{p-1}|M_T|].$$

By the Hölder inequality, we deduce that

$$\mathbb{E}[(M_J^* \wedge k)^p] \leq \frac{p}{p-1}(\mathbb{E}[(M_J^* \wedge k)^p])^{1-\frac{1}{p}}(\mathbb{E}[|M_T|^p])^{1/p}.$$

This gives $\mathbb{E}[(M_J^* \wedge k)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_T|^p$, which yields (5.14) at the limit $k \to +\infty$. $\square$

Remark 5.3 (An alternative proof, and generalization, of (5.14)). Let $\Phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ be an non-decreasing function of class C^1. Assume in a first time that $\lambda \mapsto \lambda^{-1}\Phi'(\lambda)$ is integrable around 0. We can generalize (5.15) into

$$\mathbb{E}[\Phi(M_J^* \wedge k)] = \int_0^k \Phi'(\lambda)\mathbb{P}(M_J^* > \lambda)d\lambda.$$

Using (5.10) and Fubini's theorem, we obtain the estimate

$$\mathbb{E}[\Phi(M_J^* \wedge k)] \leq \mathbb{E}\int_0^k \lambda^{-1}\Phi'(\lambda)\mathbf{1}_{\{M_J^*>\lambda\}}|M_T||d\lambda = \mathbb{E}[\Psi(M_J^* \wedge k)|M_T|], \tag{5.16}$$

where $\Psi'(\lambda) := \lambda^{-1}\Phi'(\lambda)$. We use the convexity inequality

$$sq \leq h(s) + h^*(q), \tag{5.17}$$

where h^* is the Fenchel-Legendre transform of h, defined by $h^*(q) = \sup_{s\in\mathbb{R}}(sq - h(s))$. Here we assume that h is a convex function of class C^1 with superlinear growth: $\lim_{|s|\to+\infty} |s|^{-1}|h(s)| = +\infty$. Then the sup defining $h^*(q)$ is reached at a point s_q such that $q = h'(s_q)$. By differentiating the relation $h^*(q) = s_q q - h(s_q)$, we obtain thus $\partial_q h^*(q) = s_q = (h')^{-1}(q)$. Note that if we apply (5.17) to the function $s \mapsto \theta h(s)$, where θ is a positive parameter, we have $(\theta h)^*(q) = \theta h^*(\theta^{-1}q)$ and thus

$$sq \leq \theta h(s) + \theta h^*(\theta^{-1}q). \tag{5.18}$$

Using (5.16) and (5.18) with $\theta \in (0,1)$, we see that

$$\mathbb{E}[\Phi(M_J^* \wedge k)] \leq \theta\mathbb{E}[h \circ \Psi(M_J^* \wedge k)] + \theta\mathbb{E}[h^*(\theta^{-1}|M_T|)].$$

Let us choose h such that $h \circ \Psi = \Phi$ (we will see that such an h exists and satisfies the properties assumed above). We deduce then that $\mathbb{E}[\Phi(M_J^* \wedge k)] \leq \theta(1-\theta)^{-1}\mathbb{E}[h^*(\theta^{-1}|M_T|)]$. At the limit $k \to +\infty$, this gives

$$\mathbb{E}[\Phi(M_J^*)] \leq \theta(1-\theta)^{-1}\mathbb{E}[h^*(\theta^{-1}|M_T|)].$$

By differentiation of the relation $h \circ \Psi = \theta\Phi$, we obtain $h' \circ \Psi\Psi' = \theta\Phi'$, and thus $h' \circ \Psi(\lambda) = \lambda$, which implies $\partial_q h^*(q) = \Psi(q)$. We deduce finally that

$$\mathbb{E}[\Phi(M_J^*)] \leq \frac{\theta}{1-\theta}\mathbb{E}[\Phi_2(\theta^{-1}|M_T|)], \tag{5.19}$$

where

$$\Phi_2(s) := \int_0^s \int_0^r \frac{\Phi'(\lambda)}{\lambda} d\lambda dr = \int_0^s \frac{s-\lambda}{\lambda}\Phi'(\lambda)d\lambda = s\Psi(s) - \Phi(s). \tag{5.20}$$

Equivalently to (5.19), we have

$$\mathbb{E}[\Phi(M_J^*)] \leq \frac{1}{a-1}\mathbb{E}[\Phi_2(a|M_T|)], \tag{5.21}$$

where a is a paramater in $(1, +\infty)$. The optimal value of a is obtained when

$$\mathbb{E}[\Phi_2(a|M_T|) - (a-1)|M_T|\Phi_2'(a|M_T|)] = 0.$$

Since $\Phi_2'(s) = \Psi(s)$ and $\Phi_2(s) = s\Psi(s) - \Phi(s)$, this equation is equivalent to

$$\mathbb{E}[a|M_T|\Psi(a|M_T|) - a\Phi((a|M_T|))] = 0. \tag{5.22}$$

Example 1. The power-law case. Let $p > 1$. When $\Phi(s) = s^p$, we have $\Psi(s) = \frac{p}{p-1}s^{p-1} = p's^{p-1}$, $\Psi_2(s) = \frac{1}{p-1}s^p$. Equation (5.22) takes the form $(p' - a)\mathbb{E}[|aM_T|^p] = 0$. We take thus $a = p'$ in (5.21) and obtain exactly (5.13).
Example 2. The exponential case. Consider $\Phi(s) = e^{\alpha s^2}$. We have then, by means of a change of variables,

$$\Psi(s) = 2\alpha \int_0^s e^{\alpha r^2} dr, \quad \Phi(s) = \alpha s^2 \int_0^1 r e^{\alpha s^2 r^2/2} dr, \quad \Phi_2(s) = \alpha s^2 \int_0^1 (1-r) e^{\alpha s^2 r^2/2} dr.$$

Compare the expressions of Φ and Φ_2. The integrand in $\Phi(s)$ reaches its maximum at $r = 1$. The integrand in $\Phi_2(s)$ reaches its maximum at a r_* close to 1, solution to the equation

$$1 = \alpha s^2 r_*(1 - r_*).$$

We expect therefore that $\Phi_2(s) \simeq e^{-1}\Phi(s)$ for large s (the Laplace's method should give the result - not checked). In particular, we shall have $\Phi_2 \leq C\Phi$ for a given constant $C \geq 0$. Taking $a = 2$ in (5.21), this gives

$$\mathbb{E}\left[\sup_{t\in[0,T]} e^{\alpha|M_t|^2/2}\right] \leq C\mathbb{E}\left[e^{\alpha|M_T|^2}\right], \tag{5.23}$$

where C is, possibly, a different non-negative constant.

5.1 Quadratic Variation

We always assume here that the filtration $(\mathcal{F}_t)$ satisfies the usual condition. We will study in this section the quadratic variation of a martingale. Let us first state the following result.

Theorem 5.4 (Doob-Meyer decomposition theorem)**.** *Let $(Y_t)_{t\in[0,T]}$ be a càdlàg, real-valued, bounded submartingale. Then $(Y_t)_{t\in[0,T]}$ admits a unique decomposition $Y = M + A$, where $(M_t)_{t\in[0,T]}$ is a martingale et $(A_t)_{t\in[0,T]}$ an increasing predictable process.*

Precictable processes are defined only later in Section 7.2. In the time discrete case, a process (A_n) is predictable if each A_n is $\mathcal{F}_{n-1}$-measurable, $n = 1, 2, \dots$. At time n, A_n is therefore entirely known.

Exercise 5.4. Prove Theorem 5.4 in the time discrete case.
The solution to Exercise 5.4 is here.

For processes indexed by a continuous set, we will just need the following result here[5]: *adapted processes which are a.s. continuous from the left are predictable.* The complete statement of the Doob-Meyer decomposition theorem is more general (no need to consider bounded processes in particular). Some proofs of Theorem 5.4 can be found at various places in classical textbooks on the general theory of stochastic processes. We mention also the paper [BSV12], for a recent, short, self-contained proof of the result.

Consider now $(X_t)_{t\in[0,T]}$ a càdlàg, real-valued, bounded martingale. We can apply the Doob-Meyer decomposition theorem in the particular case $Y_t = X_t^2$. What is the process A in that case? If X is a.s. continuous, then A is the quadratic variation of X, which we define below. The situation where X may have jumps is discussed in Remark 5.4.

Theorem 5.5 (Quadratic variation)**.** *Let $(X_t)_{t\in[0,T]}$ be a* continuous*, real-valued, bounded martingale. Let $\sigma = \{t_0, \dots, t_n\}$ with*

$$0 = t_0 < t_1 < \dots < t_n = T$$

be a subdivision of $[0, T]$ of size $|\sigma| = \inf_{0\le i<n}(t_{i+1} - t_i)$. We introduce $V_\sigma^{(2)}(t)$ the variation of order 2 relative to σ:

$$V_\sigma^{(2)}(t) = \sum_{i=0}^{n-1} |X_{t\wedge t_{i+1}} - X_{t\wedge t_i}|^2. \tag{5.24}$$

Then, there exists an increasing adapted continous process $t \mapsto \langle X, X\rangle_t$ such that, for all $t \in [0, T]$, $V_\sigma^{(2)}(t)$ is converging in $L^2(\Omega)$ to $\langle X, X\rangle_t$ when $|\sigma| \to 0$. The process $(\langle X, X\rangle_t)_{t\ge0}$ is called the quadratic variation *of $(X_t)_{t\ge0}$. It is the unique increasing, continuous, adapted process $(Z_t)_{t\ge0}$ such that $(X_t^2 - Z_t)$ is a martingale.*

[5] a result which is natural with regard to the time discrete case, see [Dur84, p. 49]

Sometimes the notation $\langle X\rangle_t$ is used to denote the quadratic variation. We will insist on using the notation $\langle X,X\rangle_t$ however, since this indicates that the object is quadratic in X. It is also consistent with the definition of the cross-quadratic variation of two martingales X and Y which is defined by polarization:

$$\langle X,Y\rangle_t = \frac{1}{4}\left[\langle X+Y,X+Y\rangle_t - \langle X-Y,X-Y\rangle_t\right].$$

Proof of Theorem 5.5. The difference of two non-increasing functions is a function with bounded variation. Uniqueness in the statement of Theorem 5.5 then comes from the fact that a martingale with a.s. bounded variation is constant, [RY99, Proposition IV-1.2]. Uniqueness also comes from the uniqueness statement in Theorem 5.4 since an adapted continuous process is predictable.
To prove the convergence of $V_\sigma^{(2)}(t)$ in $L^2(\Omega)$, we consider a sequence of subdivisions σ^m with $|\sigma^m|\to 0$. We want to show a Cauchy condition of the type

$$\mathbb{E}|V_{\sigma^p}^{(2)}(t) - V_{\sigma^q}^{(2)}(t)|^2 \to 0, \tag{5.25}$$

when $p,q\to+\infty$. From σ^p and σ^q we can form a refined subdivision $\sigma^{p,q}$ common to σ^p and σ^q by taking all the points of both subdivisions. Using $V_{\sigma^{p,q}}^{(2)}(t)$ as a common element of comparison, we see that is sufficient to show that

$$\lim_{|\sigma|\to 0}\mathbb{E}|V_\sigma^{(2)}(t) - V_{\sigma'}^{(2)}(t)|^2 = 0, \tag{5.26}$$

where σ' is a refinement of σ. We use the following reduction and notations:

1. we assume without loss of generality that $X(0)=0$. Let $M>0$ be such that $|X_t|\le M$ for all $t\in[0,T]$,
2. subdivisions: $\sigma'=\{t_k\}$ is the fine one, $\sigma=\{s_l\}$ the coarsest one,
3. the final indices relatively to t are respectively
$$K=\sup\{k;t_k\le t\},\quad L=\sup\{l;s_l\le t\},$$
4. increments: fine ones: $\zeta(t_k)=X(t_k)-X(t_{k-1})$, big ones: $Z(s_l)=X(s_l)-X(s_{l-1})$, intermediary ones: $z(t_k)=X(t_k)-X(\pi t_k)$ where
$$\pi t_k := \max\{s_l; s_l\le t_k\}$$
(the action of π is to project a $t_k\in\sigma'$ onto the closest element of σ below t_k).

Note that

$$Z(s_l) = \sum_{\{k:\pi t_{k-1}=s_{l-1}\}} \zeta(t_k). \tag{5.27}$$

In particular, due to (5.27), the difference $B = \sum_{l \le L} |Z(s_l)|^2 - \sum_{k \le K} |\zeta(t_k)|^2$ is

$$B = \sum_l \left[\left| \sum_{\{k : \pi t_{k-1} = s_{l-1}\}} \zeta(t_k) \right|^2 - \sum_{\{k : \pi t_{k-1} = s_{l-1}\}} |\zeta(t_k)|^2 \right],$$

and, by developing the square, this gives

$$B = 2 \sum_l \sum_{\{k : \pi t_{k-1} = s_{l-1}\}} \sum_{\{j < k : \pi t_{j-1} = s_{l-1}\}} \zeta(t_k)\zeta(t_j) = 2 \sum_l \sum_{\{k : \pi t_{k-1} = s_{l-1}\}} \zeta(t_k) z(t_{k-1}),$$

whence

$$B = 2 \sum_k \xi(t_k), \quad \xi(t_k) := \zeta(t_k) z(t_{k-1}).$$

We have $\mathbb{E}\left[\xi(t_{k+1}) | \mathcal{F}_{t_k}\right] = 0$, which implies

$$\frac{1}{4}\mathbb{E}|B|^2 = \sum_{k \le K} \mathbb{E}|\xi(t_k)|^2 + 2 \sum_{j < k \le K} \mathbb{E}[\xi(t_j)\xi(t_k)] = \sum_{k \le K} \mathbb{E}|\xi(t_k)|^2 \tag{5.28}$$

since $\mathbb{E}[\xi(t_j)\xi(t_k)] = \mathbb{E}(\mathbb{E}[\xi(t_j)\xi(t_k)|\mathcal{F}_{t_j}]) = \mathbb{E}(\xi(t_j)\mathbb{E}[\xi(t_k)|\mathcal{F}_{t_j}]) = 0$ for $j < k$. Let $\varepsilon > 0$. Let

$$\tau_l = \min\{t_k \ge s_l ; |z(t_k)| > \varepsilon\} \cup \{s_{l+1}\}, \quad \beta_l = s_{l+1}. \tag{5.29}$$

Note that each τ_l is a stopping time with respect to $(\mathcal{F}_t)$ since deciding the occurrence of the event $\{\tau_l \le t\}$ is something non-trivial only when $t \in [s_l, s_{l+1}]$ and, in that case, this only requires to know $M(s)$ up to the time t_k such that $t_k \le t < t_{k+1}$. We use the following estimate on (5.28): $\mathbb{E}|B|^2$ is bounded by the sum of

$$4\mathbb{E} \sum_l \sum_{\{\pi t_{k-1} \le t_k < \tau_{l-1}\}} |\xi(t_k)|^2 \tag{5.30}$$

and

$$4\mathbb{E} \sum_l \sum_{\{\tau_{l-1} \le t_{k-1} < t_k \le s_l\}} \mathbf{1}_{\tau_{l-1} < s_l} |\xi(t_k)|^2. \tag{5.31}$$

The first term (5.30) is bounded by

$$4\varepsilon^2 \mathbb{E} \sum_k |\zeta(t_k)|^2 = 4\varepsilon^2 \mathbb{E}|X(t_K)|^2 \le 4M^2\varepsilon^2. \tag{5.32}$$

The equality in (5.32) comes from the decomposition

$$|\zeta(t_k)|^2 = |X(t_k) - X(t_{k-1})|^2 = |X(t_k)|^2 - |X(t_{k-1})|^2 - 2X_{t_{k-1}}\zeta(t_k) \tag{5.33}$$

and the martingale property $\mathbb{E}[\zeta(t_k)|\mathcal{F}_{t_{k-1}}] = 0$. In the second term (5.31), we estimate the intermediary increment $z(t_k)$ by $2M$, which gives the bound by

$$16M^2\mathbb{E}\sum_l \sum_{\{\tau_{l-1}\leq t_{k-1}<t_k\leq s_l\}} \mathbf{1}_{\tau_{l-1}<s_l}|\zeta(t_k)|^2$$
$$= 16M^2\mathbb{E}\sum_l \mathbf{1}_{\tau_{l-1}<s_l} \sum_{\{\tau_{l-1}\leq t_{k-1}<t_k\leq s_l\}} \mathbb{E}[|\zeta(t_k)|^2|\mathcal{F}_{\tau_{l-1}}].$$

By (5.33) and the martingale property, we have

$$\sum_{\{\tau_{l-1}\leq t_{k-1}<t_k\leq s_l\}} \mathbb{E}[|\zeta(t_k)|^2|\mathcal{F}_{\tau_{l-1}}] = \mathbb{E}[|X(s_l)|^2|\mathcal{F}_{\tau_{l-1}}] - |X(\tau_{l-1})|^2. \tag{5.34}$$

To obtain (5.34), it is sufficient to write $\tau_{l-1} = \tau_{l-1}\wedge t_{k-1}$ if $\tau_{l-1\leq t_{k-1}}$ and thus

$$\sum_{\{\tau_{l-1}\leq t_{k-1}<t_k\leq s_l\}} \mathbb{E}[|\zeta(t_k)|^2|\mathcal{F}_{\tau_{l-1}\wedge t_{k-1}}] = \mathbb{E}\left[\sum_{\{\tau_{l-1}\leq t_{k-1}<t_k\leq s_l\}} |\zeta(t_k)|^2 \middle| \mathcal{F}_{\tau_{l-1}\wedge t_{k-1}}\right]$$
$$= \mathbb{E}\left[\sum_{\{\tau_{l-1}\leq t_{k-1}<t_k\leq s_l\}} \mathbb{E}[|\zeta(t_k)|^2|\mathcal{F}_{t_{k-1}}] \middle| \mathcal{F}_{\tau_{l-1}}\right]. \tag{5.35}$$

Then we use (5.33). Note that (5.35) is satisfied since $\mathcal{F}_{\tau_{l-1}\wedge t_{k-1}} \subset \mathcal{F}_{t_{k-1}}$ (*cf.* Question 2 of Exercise 4.13). We deduce from (5.34) that (5.31) is bounded by

$$16M^2\mathbb{E}\sum_l \mathbf{1}_{\tau_{l-1}<s_l}(\mathbb{E}[|X(s_l)|^2|\mathcal{F}_{\tau_{l-1}}] - |X(\tau_{l-1})|^2) \tag{5.36}$$

Let $N > 0$ (that will be large). Let γ_N denote the stopping time

$$\gamma_N = \min\left\{s_l; \sum_{i=1}^l \mathbf{1}_{\tau_{i-1}<s_i} = N\right\} \cup \{s_L\}.$$

Let λ_N be the corresponding index: $\gamma_N = s_{\lambda_N}$. In (5.36), we consider the sum over the indices $\{\lambda_N < l \leq L\}$. Using the simple estimate $\mathbf{1}_{\tau_{l-1}<s_l} \leq 1$, the fact that (consequence of Lemma 5.1)

$$\mathbb{E}[|X(s_l)|^2|\mathcal{F}_{\tau_{l-1}}] - |X(\tau_{l-1})|^2 \geq 0 \tag{5.37}$$

and the identity $\mathbb{E}\left(\mathbb{E}[|X(s_l)|^2|\mathcal{F}_{\tau_{l-1}}]\right) = \mathbb{E}|X(s_l)|^2$, we obtain a telescopic sum and, thus, a bound by

$$16M^2\mathbb{E}\left[|X(s_L)|^2 - |X(\gamma_N)|^2\right] \leq 16M^4\mathbb{P}(\lambda_N < L). \tag{5.38}$$

Let η (random, a.s. positive) be a modulus of uniform continuity of $t \mapsto X(t)$ associated to ε. Let $|\sigma| = \max(s_{l+1} - s_l)$. We have $\mathbf{1}_{\tau_{l-1}<s_l} \leq \mathbf{1}_{\eta\leq|\sigma|}$. Therefore, using the bound by M on $X(t)$, the sum over the indices $\{l \leq \lambda_N\}$ in (5.36) can be bounded by

$$16M^4N\mathbb{P}(\eta \leq |\sigma|). \tag{5.39}$$

Gathering the estimates (5.32), (5.38), (5.39), we conclude that

$$\mathbb{E}|B|^2 \leq 4M^2\varepsilon^2 + 16M^4\mathbb{P}(\gamma_N < s_L) + 16M^4 N\mathbb{P}(\eta \leq |\sigma|). \tag{5.40}$$

Since $\lim_{N\to+\infty}\mathbb{P}(\gamma_N < s_L) = 0$ and $\lim_{|\sigma|\to 0}\mathbb{P}(\eta \leq |\sigma|) = 0$, choosing N large, then $|\sigma|$ small gives $\mathbb{E}|B|^2 < \varepsilon$. This proves the convergence (5.26). Note that B, which depends on t actually, is a continuous martingale since $V_\sigma^{(2)}(t)$ is continuous in t. By using the Doob's inequality (5.13), our previous considerations applied at the final time $t = T$ gives the Cauchy condition

$$\mathbb{E}\sup_{t\in[0,T]}|V_{\sigma^p}^{(2)}(t) - V_{\sigma^q}^{(2)}(t)|^2 \to 0, \tag{5.41}$$

which is stronger than (5.41), and which shows that $(\langle X, X\rangle_t)$ is continuous. There remains to prove that $(M(t))$, where $M(t) = X_t^2 - \langle X, X\rangle_t$, is a martingale. Let $M_\sigma(t) = X_t^2 - V_\sigma^{(2)}(t)$. Let $0 \leq s < t$, and $t_{n+1} = \min\{t_i; t_i \geq t\}$, $t_{l+1} = \min\{t_i; t_i \geq s\}$. We may assume $t_n \geq s$. We have the expansion

$$M_\sigma(t) = X_t^2 - \sum_{i=0}^{n-1}|X_{t_{i+1}} - X_{t_i}|^2 - |X_t - X_{t_n}|^2,$$

which gives the identity

$$M_\sigma(t) - M_\sigma(s) = (X_t^2 - X_s^2) - \sum_{i=l}^{n-1}|X_{t_{i+1}} - X_{t_i}|^2 - |X_t - X_{t_n}|^2 + |X_s - X_{t_l}|^2.$$

Since $\mathbb{E}\left[|X_\sigma - X_r|^2|\mathcal{F}_s\right] = \mathbb{E}\left[X_\sigma^2 - X_r^2|\mathcal{F}_s\right]$ if $r, \sigma \geq s$, we obtain the identity

$$\mathbb{E}\left[M_\sigma(t) - M_\sigma(s)|\mathcal{F}_s\right] = \mathbb{E}\left[|X_s - X_{t_l}|^2 - |X_{t_{l+1}} - X_{t_l}|^2|\mathcal{F}_s\right]. \tag{5.42}$$

by conditioning with respect to $\mathcal{F}_s$. The right-hand side of (5.42) tends to 0 when $|\sigma| \to 0$ by continuity (and boundedness) of the process (X_t). Taking the limit $|\sigma| \to 0$ in (5.42) gives thus the desired result $\mathbb{E}\left[M(t) - M(s)|\mathcal{F}_s\right] = 0$. □

Exercise 5.5. Give the quadratic variation of the one-dimensional Wiener process. *The solution to Exercise 5.5 is here.*

Remark 5.4. Consider the case where $X = N$, a Poisson Process of intensity λ. It is quite clear that, when $|\sigma| \to 0$, the sum of the increments $V_\sigma^{(2)}(t)$ should converge (at least if t is not a time of jump) to the sum of the square of all the jumps that have occurred before time t, *i.e.* $\sum_{s\leq t}(\Delta X_s)^2$, where, for a general càdlàg process, we set $\Delta X_t := X_t - X_{t-}$. In the case where X is the Poisson Process, the jumps have size 1, hence $(\Delta X_s)^2 = \Delta X_s$, and we find (quite informally) that the quadratic variation of the Poisson process N is N itself:

$$[N, N]_t = N_t.$$

Here we have used the notation $[\cdot,\cdot]_t$ for the quadratic variation, the notation $\langle\cdot,\cdot\rangle_t$ is used to denote the predictable process as the A of Theorem 5.4 (compensator), see the final lines of this remark. For a general semimartingale[6], the quadratic variation is defined by

$$[X,X]_t = X_t^2 - X_0^2 - 2\int_0^t X_{s-}dX_s. \tag{5.43}$$

We do not explain the meaning of the stochastic integral in (5.43) either, but you may guess that, in the case where (X_t) is a mere jump process, only jumps with their respective size should contribute to the stochastic integral, giving thus

$$\begin{aligned}[X,X]_t &= X_t^2 - X_0^2 - 2\sum_{s\le t} X_{s-}\Delta X_s = \sum_{s\le t}\Delta(X_s)^2 - 2\sum_{s\le t}X_{s-}\Delta X_s\\ &= \sum_{s\le t}(X_s + X_{s-} - 2X_{s-})\Delta X_s = \sum_{s\le t}(\Delta X_s)^2.\end{aligned}$$

This computation, done in the case where (X_t) is a mere jump process, shows that the definition (5.43) seems consistent with the approach by sum of square of increments in Theorem 5.5 (and the reason for this is the fact that the stochastic integral is defined first for step-functions, using increments). Besides, since the stochastic integral is a martingale, $[X,X]_t$ in (5.43) is such that $X_t^2 - [X,X]_t$ is a martingale. However, in the case of càdlàg processes, $t \mapsto [X,X]_t$ is not the only process that realizes this property. Indeed, if we come back to the case of the Poisson Process, for example, we know by independence of the increments that $(N_t - \lambda t)$ is a martingale. Since $[N,N]_t = N_t$, we deduce that the process $A_t = \lambda t$ is also such that $N_t^2 - A_t$ is a martingale. What is the difference between (A_t) and $([N,N]_t)$? We notice that (A_t) is predictable (it is deterministic), therefore it fulfills the condition of Theorem 5.4 applied to $Y = N^2$, while $[N,N]$ does not (it is increasing, but certainly not predictable, being a jump process). For a general càdlàg process X, the notation $[X,X]_t$ will denote the quadratic variation as defined in (5.43), while the notation $\langle X,X\rangle$ will denote the process A given by Theorem 5.4 applied to $Y = X^2$. Sometimes, $\langle X,X\rangle_t$ is called the predictable quadratic variation. In terms of compensator, [JS03, p. 32], $\langle X,X\rangle$ is the predictable compensator of $[X,X]$. This immediately follows from the definition of predictable compensator, since $[X,X] - \langle X,X\rangle$ is a martingale.

Proposition 5.6 (Quadratic variation)**.** *Under the hypotheses of Theorem 5.5, let*

$$\bar{V}_\sigma^{(2)}(t) = \sum_{i=0}^{n-1}\mathbb{E}\left[|X_{t\wedge t_{i+1}} - X_{t\wedge t_i}|^2|\mathcal{F}_{t_i}\right]. \tag{5.44}$$

Then $\bar{V}_\sigma^{(2)}(t)$ is converging in $L^2(\Omega)$ to $\langle X,X\rangle_t$ when $|\sigma|\to 0$.

[6] we do not explain the terms here, see [JS03], p. 32 for the definition of compensator, and p. 42 for the definition of semimartingale, but let us note state that a martingale is a semimartingale

Proof of Proposition 5.6. Let $D = V_\sigma^{(2)}(t) - \bar{V}_\sigma^{(2)}(t)$. Using the notations of the proof of Theorem 5.5, we have

$$D = \sum_{k \le K} \theta(t_k), \quad \theta(t_k) := |\zeta(t_k)|^2 - \mathbb{E}[|\zeta(t_k)|^2 | \mathcal{F}_{t_{k-1}}].$$

Since $\mathbb{E}[\theta(t_k)|\mathcal{F}_{t_{k-1}}] = 0$, we obtain, as in (5.28),

$$\mathbb{E}|D|^2 = \sum_k \mathbb{E}|\theta(k)|^2 \le 4 \sum_k \mathbb{E}|\zeta(t_k)|^4.$$

By adapting the proof Theorem 5.5, we can show then that $\mathbb{E}|D|^2 \to 0$ when $|\sigma| \to 0$. □

Exercise 5.6. Give the details of the end of the proof of Proposition 5.6. *The solution to Exercise 5.6 is here.*

Exercise 5.7. Assume that (X_t) is a jump process as in Exercise 4.11. Suppose that, for $n \ge 1$, each X_n is drawn independently on X_{n-1} according to a law ν. Try to guess what would be the limit of $\bar{V}_\sigma^{(2)}(t)$ in that case. *The solution to Exercise 5.7 is here.*

We end this section with the following result (5.45), which is a particular case of the more general inequality of Burkholder, Davis, Gundy [BDG72], [Bau14, Theorem 5.70].

Proposition 5.7. *Let $p \ge 2$. There exists a constant $C_{\mathrm{BDG}}(p) \ge 0$ such that, for all continuous, real-valued martingale $(M_t)_{t\in[0,T]}$ such that $\mathbb{E}|M_T|^p < +\infty$ and $M_0 = 0$, the inequality*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |M_t|^p\right] \le C_{\mathrm{BDG}}(p)\mathbb{E}\left[\langle M, M\rangle_T^{p/2}\right] \tag{5.45}$$

is satisfied.

Remark 5.5. The result is true for càdlàg martingales, with $\langle M, M\rangle_T$ replaced with $[M, M]_T$ (*cf.* Remark 5.4). See, for example, [MR14], for a proof "using almost only stochastic calculus".

(Partial) proof of Proposition 5.7. By the Doob's inequality (5.13), we have

$$\mathbb{E}\left[\sup_{t\in[0,T]} |M_t|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_T|^p. \tag{5.46}$$

Under the hypotheses of Proposition 5.7, we also have

$$\mathbb{E}\left[\langle M, M\rangle_T^{p/2}\right] \le \tilde{C}_{\mathrm{BDG}}(p)\mathbb{E}\left[\sup_{t\in[0,T]} |M_t|^p\right].$$

This justifies in particular that $\mathbb{E}\left[\langle M, M\rangle_T^{p/2}\right]$ is finite when $\mathbb{E}|M_T|^p < +\infty$. We will admit this fact here to do our proof. Let $\varphi(s) = |s|^p$. Let $\sigma = \{0 = t_0 < \cdots < t_N = T\}$

be a subdivision of the interval $[0,T]$. We decompose $\varphi(M_T)$ as the sum of the increments $\varphi(M_{t_{i+1}})-\varphi(M_{t_i})$. Using the Taylor formula and the fact that $\varphi''\colon s\mapsto p(p-1)|s|^{p-2}$ is increasing on $\mathbb{R}_+$, we have, a.s.

$$\varphi(M_{t_{i+1}})-\varphi(M_{t_i})\leq\varphi'(M_{t_i})(M_{t_{i+1}}-M_{t_i})+\frac{1}{2}\varphi''(M_T^*)|M_{t_{i+1}}-M_{t_i}|^2,$$

where $M_T^*=\sup_{t\in[0,T]}|M_t|$. Taking the conditional expectation with respect to $\mathcal{F}_{t_i}$, summing on i and taking expectation, we obtain

$$\mathbb{E}\varphi(M_T)\leq\frac{1}{2}\mathbb{E}\left[\varphi''(M_T^*)\bar{V}_\sigma^{(2)}\right],$$

where $\bar{V}_\sigma^{(2)}$ is defined by (5.44). This gives, using (5.46),

$$\mathbb{E}\left[(M_T^*)^p\right]\leq\left(\frac{p}{p-1}\right)^p\frac{p(p-1)}{2}\mathbb{E}\left[(M_T^*)^{p-2}\bar{V}_\sigma^{(2)}\right].$$

By the Hölder inequality, we deduce that

$$\mathbb{E}\left[(M_T^*)^p\right]\leq\left(\frac{p}{p-1}\right)^p\frac{p(p-1)}{2}\mathbb{E}\left[(M_T^*)^p\right]^{\frac{p-2}{p}}\mathbb{E}\left[|\bar{V}_\sigma^{(2)}|^{p/2}\right]^{2/p}.$$

At the limit $|\sigma|\to 0$, we obtain (5.45) with the constant

$$C_{\mathrm{BDG}}(p)=\left[\left(\frac{p}{p-1}\right)^p\frac{p(p-1)}{2}\right]^{p/2}.$$

□

5.2 Law and paths of a Markov process

The two results Theorem 5.8 (“Dynkin's formula”, [Pro05, p. 56]) and Theorem 5.10 below give a martingale characterization of a Markov process. This characterization requires the knowledge of the generator $\mathcal{L}$. This is the reason why we will put so much emphasis on the generator in the diffusion-approximation results of Section 6. To establish these results, we will also need Proposition 5.9, which completes Theorem 5.8.

Theorem 5.8. *Let E be a separable Banach space, let $(\mathcal{F}_t)$ be a filtration. Let (X_t) be an E-valued time-homogeneous Markov process with respect to $(\mathcal{F}_t)$, with semi-group of transition operators $(P_t)_{t\geq 0}$ satisfying (4.16). Let $\mathcal{L}$ be the generator of (P_t). Assume that (X_t) is a.s. continuous. Then, for all φ in the domain of $\mathcal{L}$,*

$$M_\varphi(t):=\varphi(X_t)-\varphi(X_0)-\int_0^t\mathcal{L}\varphi(X_s)ds \tag{5.47}$$

is a $(\mathcal{F}_t)$-martingale.

Remark 5.6. Up to a modification, we can assume that $(\omega, t) \mapsto X_t(\omega)$ is measurable. Consequently the integral in (5.47) is, at fixed time t, a random variable. Note also that (X_t) is $(\mathcal{F}_t)$-adapted since the Markov identity $\mathbb{E}[\varphi(X_{t+s})|\mathcal{F}_t] = P_s\varphi(X_t)$ gives $\varphi(X_t) = \mathbb{E}[\varphi(X_t)|\mathcal{F}_t]$ when $s = 0$. Consequently, all the terms in (5.47) are $\mathcal{F}_t$-measurable.

Proposition 5.9. *Under the hypotheses of Theorem 5.8, assume furthermore that $|\varphi|^2$ is in the domain of $\mathcal{L}$. Then the quadratic variation of the càdlàg martingale $(M_\varphi(t))$ defined by (5.47) is given by*

$$\langle M_\varphi, M_\varphi\rangle_t = \int_0^t (\mathcal{L}|\varphi|^2 - 2\varphi\mathcal{L}\varphi)(X_s)ds, \tag{5.48}$$

for all $t \geq 0$.

Theorem 5.10. *Let E be a separable Banach space. Let $(P_t)_{t\geq 0}$ be a π-contraction semi-group (Definition 4.5) such that $t \mapsto P_t\varphi(x)$ is continuous for all $\varphi \in C_b(E)$, for all $x \in E$. Let $\mathcal{L}$, with domain $\mathcal{D}(\mathcal{L})$, be the infinitesimal generator associated to $(P_t)_{t\geq 0}$. Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration, let $(X_t)_{t\geq 0}$ be an E-valued process such that $(\omega, t) \mapsto X_t(\omega)$ is measurable $\Omega \times \mathbb{R}_+ \to E$. Assume that, for all $\varphi \in \mathcal{D}(\mathcal{L})$, the process $(M_\varphi(t))_{t\geq 0}$ defined by (5.47) is an $(\mathcal{F}_t)$-martingale. Then $(X_t)_{t\geq 0}$ satisfies*

$$\mathbb{E}\Big[\varphi(X_{s+t})\Big|\mathcal{F}_t\Big] = P_s\varphi(X_t), \tag{5.49}$$

for all $\varphi \in C_b(E)$, for all $t, s \geq 0$.

Proof of Theorem 5.8 and Proposition 5.9. Let $0 \leq s \leq t$. By Remark 5.6, we have

$$\mathbb{E}[M_\varphi(t)|\mathcal{F}_s] - M_\varphi(s) = \mathbb{E}[M_\varphi(t) - M_\varphi(s)|\mathcal{F}_s] = P_{t-s}\varphi(X_s) - \varphi(X_s) - \int_s^t [P_{\sigma-s}\mathcal{L}\varphi](X_s)d\sigma.$$

We use the relation $\frac{d}{dt}P_t\varphi(x) = P_t\mathcal{L}\varphi(x)$ (see (4.17)) to obtain the martingale property. Indeed, this gives

$$P_{t-s}\varphi - \varphi = \int_s^t P_{\sigma-s}\mathcal{L}\varphi d\sigma,$$

and thus $\mathbb{E}[M_\varphi(t)|\mathcal{F}_s] - M_\varphi(s) = 0$. The proof of (5.48) is divided in several steps. By $C(\varphi)$, we will denote any constant that depend on φ and may vary from lines to lines. We fix a subdivision $\sigma = (t_i)_{0,n}$ of $[0, T]$ and introduce the notation

$$A_t = \int_0^t (\mathcal{L}|\varphi|^2 - 2\varphi\mathcal{L}\varphi)(X_s)ds. \tag{5.50}$$

In a first step, we show that

$$A_t = \lim_{|\sigma|\to 0} \sum_{i=0}^{n-1} \mathbb{E}\left[A_{t\wedge t_{i+1}} - A_{t\wedge t_i}|\mathcal{F}_{t_i}\right]. \tag{5.51}$$

Indeed, we have

$$A_t = \sum_{i=0}^{n-1} A_{t\wedge t_{i+1}} - A_{t\wedge t_i}, \tag{5.52}$$

and $\zeta(t_{i+1}) := A_{t\wedge t_{i+1}} - A_{t\wedge t_i} - \mathbb{E}\left[A_{t\wedge t_{i+1}} - A_{t\wedge t_i} | \mathcal{F}_{t_i}\right]$ satisfies

$$\mathbb{E}[\zeta(t_i)\zeta(t_j)] = 0, \quad i \neq j, \quad |\zeta(t_{i+1})| \leq C(\varphi)(t_{i+1} - t_i), \tag{5.53}$$

where $C(\varphi) = \|\mathcal{L}\varphi^2\|_{BM(E)} + 2\|\varphi\|_{BM(E)}\|\mathcal{L}\varphi\|_{BM(E)}$. It follows that

$$\mathbb{E}\left|\sum_{i=0}^{n-1} \zeta(t_{i+1})\right|^2 = \mathbb{E}\sum_{i=0}^{n-1} |\zeta(t_{i+1})|^2 \leq C(\varphi)T|\sigma|,$$

which tends to 0 when $|\sigma| \to 0$. Using (5.52), we obtain (5.51). In a second step we prove that

$$|M_\varphi(t_{i+1}) - M_\varphi(t_i)|^2 = |\varphi(X_{t_{i+1}}) - \varphi(X_{t_i})|^2 + R_{t_i,t_{i+1}}, \tag{5.54}$$

with

$$\mathbb{E}\sum_{i=0}^{n-1} |R_{t_i,t_{i+1}}| = \mathcal{O}(|\sigma|^{1/2}). \tag{5.55}$$

By definition of $M_\varphi(t)$, (5.54) is satisfied with a remainder term

$$R_{t_i,t_{i+1}} = \left|\int_{t_i}^{t_{i+1}} \mathcal{L}\varphi(X_s)ds\right|^2 - 2(\varphi(X_{t_{i+1}}) - \varphi(X_{t_i}))\int_{t_i}^{t_{i+1}} \mathcal{L}\varphi(X_s)ds. \tag{5.56}$$

Using the fact that $\varphi^2 \in \mathcal{D}(\mathcal{L})$, we have also

$$\begin{aligned} |\varphi(X_{t_{i+1}}) - \varphi(X_{t_i})|^2 &= M_{\varphi^2}(t_{i+1}) - M_{\varphi^2}(t_i) - 2\varphi(X_{t_i})(M_\varphi(t_{i+1}) - M_\varphi(t_i)) \\ &\quad + \int_{t_i}^{t_{i+1}} \mathcal{L}\varphi^2(X_s)ds - 2\varphi(X_{t_i})\int_{t_i}^{t_{i+1}} \mathcal{L}\varphi(X_s)ds. \end{aligned}$$

It follows that

$$\mathbb{E}[|\varphi(X_{t_{i+1}}) - \varphi(X_{t_i})|^2 | \mathcal{F}_{t_i}] = \int_{t_i}^{t_{i+1}} \mathbb{E}\left[\left(\mathcal{L}\varphi^2(X_s) - 2\varphi(X_{t_i})\mathcal{L}\varphi(X_s)\right) | \mathcal{F}_{t_i}\right] ds. \tag{5.57}$$

Taking expectation in (5.57), we get the following bound.

$$\mathbb{E}[|\varphi(X_{t_{i+1}}) - \varphi(X_{t_i})|^2] \leq C_\varphi(t_{i+1} - t_i). \tag{5.58}$$

Consider now the cross-product term in the right-hand side of (5.56). Using Young's inequality with a parameter $\eta > 0$, we see that the term $\mathbb{E}|R_{t_i,t_{i+1}}|$ can be bounded by

$$(1 + \eta^{-1})\mathbb{E}\left|\int_{t_i}^{t_{i+1}} \mathcal{L}\varphi(X_s)ds\right|^2 + \eta\mathbb{E}[|\varphi(X_{t_{i+1}}) - \varphi(X_{t_i})|^2],$$

and thus, taking $\eta = (t_{i+1}-t_i)^{1/2}$, bounded from above by $C(\varphi)(t_{i+1}-t_i)^{3/2}$. This gives (5.55). We conclude in a third step. Using the characterization in Proposition 5.6 of the quadratic variation and (5.51), (5.55), (5.57), we see that

$$\langle M_\varphi, M_\varphi\rangle_t = A_t + \varepsilon(|\sigma|) + r(t,\sigma), \tag{5.59}$$

where $\varepsilon(|\sigma|) \to 0$ in $L^2(\Omega)$ when $|\sigma| \to 0$ and

$$|r(t,\sigma)| \leq 2\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} |(\varphi(X_{t_i}) - \varphi(X_s))\mathcal{L}\varphi(X_s)|\, ds.$$

We have

$$|r(t,\sigma)| \leq C(\varphi)\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} |\varphi(X_{t_i}) - \varphi(X_s)|\, ds$$

and an estimate similar to (5.58) (obtained by working on the increment $\varphi(X_s)-\varphi(X_{t_i})$ instead of $\varphi(X_{t_{i+1}})-\varphi(X_{t_i})$) shows that $\mathbb{E}|\varphi(X_s)-\varphi(X_{t_i})|^2 \leq C(\varphi)(s-t_i)$. We deduce that $r(t,\sigma)$ is also converging to 0 in $L^2(\Omega)$ when $|\sigma| \to 0$. $\square$

Remark 5.7 (The càdlàg case). The proof of Theorem 5.8-Proposition 5.9 does not use the continuity of the trajectories $t \mapsto X_t$ (we simply use the continuity in quadratic mean (5.58)). Let us replace the hypothesis that (X_t) is a.s. continuous by the assumption that it is a.s. càdlàg. Inspecting the proof of Theorem 5.8-Proposition 5.9, we see that we obtain the following result: $M_\varphi(t)$ in (5.47) is a $(\mathcal{F}_t)$-martingale, and the variation

$$\bar{V}_\sigma^{(2)}(t) = \sum_{i=0}^{n-1}\mathbb{E}\left[|M_\varphi(t\wedge t_{i+1}) - M_\varphi(t\wedge t_i)|^2|\mathcal{F}_{t_i}\right]$$

(*cf.* (5.44)) is converging in $L^2(\Omega)$ to the continuous process A_t defined by (5.50). If we inspect now the end of the proof of Theorem 5.5 and use additionally the tower property (2.44), we see (compare to (5.42)) that, for $0 \leq s < t$ and $Z_\sigma(t) := |M_\varphi(t)|^2 - \bar{V}_\sigma^{(2)}(t)$,

$$\mathbb{E}\left[Z\sigma(t) - Z_\sigma(s)|^2|\mathcal{F}_s\right] = \mathbb{E}\left[\mathbb{E}\left[|M_\varphi(s)|^2 - |M_\varphi(t_{l+1})|^2|\mathcal{F}_{t_l}\right]|\mathcal{F}_s\right]$$

By (5.58), we deduce, at the limit $|\sigma| \to 0$, that $|M_\varphi(t)|^2 - A_t$ is a martingale. Since $t \mapsto A_t$ is predictable (continuous and adapted), this establishes the following fact: *Proposition* 5.9 *holds true in the càdlàg case, where* $\langle M_\varphi, M_\varphi\rangle_t$ *denotes the predictable quadratic variation* (*cf.* Remark 5.4).

Remark 5.8 (Time dependent test-functions). It is easy to adapt the results of Theorem 5.8 and Proposition 5.9 to the case where the test-function depends on t also. Under adequate hypotheses on ψ, the stochastic process

$$M_\psi(t) := \psi(t,X_t) - \psi(0,X_0) - \int_0^t [(\partial_s + \mathcal{L})\psi](s,X_s)ds \tag{5.60}$$

is a martingale with quadratic variation

$$\langle M_\psi, M_\psi \rangle_t = \int_0^t [(\partial_s + \mathcal{L})\psi^2 - 2\psi(\partial_s + \mathcal{L})\psi](s, X_s) ds. \tag{5.61}$$

See (5.62) for example.

Proof of Theorem 5.10. Let $\varphi \in \mathcal{D}(\mathcal{L})$. Let us first show that the martingale property for (5.47) implies the martingale property for (5.60) with $\psi(t,x) = \theta(t)\varphi(x)$, $\theta \in C^1(\mathbb{R}_+)$. If $(M_t)_{t\geq 0}$ is a continuous martingale, then

$$t \mapsto M_t\theta(t) - \int_0^t M_\sigma \theta'(\sigma) d\sigma$$

is a martingale. For M_t given by (5.47), using the Fubini theorem we obtain (5.60). Taking now $\theta(t) = e^{-\lambda t}$, $\lambda > 0$ gives us

$$e^{-\lambda(t+T)}\mathbb{E}[\varphi(X_{t+T})|\mathcal{F}_t] = e^{-\lambda t}\varphi(X_t) + \mathbb{E}\left[\int_t^{t+T} \lambda e^{-\lambda s}(\lambda^{-1}\mathcal{L} - \mathrm{Id})\varphi(X_s) ds \middle| \mathcal{F}_t\right]. \tag{5.62}$$

Doing the change of variable $s = s' + t$ in the integral shows that

$$\varphi(X_t) = e^{-\lambda T}\mathbb{E}[\varphi(X_{t+T})|\mathcal{F}_t] - \mathbb{E}\left[\int_0^T \lambda e^{-\lambda s}(\lambda^{-1}\mathcal{L} - \mathrm{Id})\varphi(X_{s+t}) ds \middle| \mathcal{F}_t\right].$$

We let $T \to +\infty$ to obtain

$$\varphi(X_t) = \mathbb{E}\left[\int_0^{+\infty} \lambda e^{-\lambda s}(\mathrm{Id} - \lambda^{-1}\mathcal{L})\varphi(X_{s+t}) ds \middle| \mathcal{F}_t\right]. \tag{5.63}$$

The convergence is easy to justify since φ and $\mathcal{L}\varphi$ are bounded. Compare (5.63) to Formula (4.24) for the resolvent. Actually, both (4.24) and (5.63) can be written more concisely by introducing a random variable independent τ with exponential distribution of parameter λ. We may work on the probability space $(\Omega, \mathcal{F})$ (it suffices to assume independence of τ and $(\mathcal{F}_t)_{t\geq 0}$). However, the lines below will be more explicit if we consider that τ is defined on a probability space $(\Omega^\sharp, \mathcal{F}^\sharp, \mathbb{P}^\sharp)$. Let $J_\lambda\varphi := \lambda R_\lambda \varphi$. We rewrite (4.24) and (5.63) as

$$J_\lambda \varphi = \mathbb{E}^\sharp P_\tau \varphi, \quad \varphi(X_t) = \mathbb{E}^\sharp \mathbb{E}\left[(J_\lambda^{-1}\varphi)(X_{\tau+t}) \middle| \mathcal{F}_t\right],$$

respectively. By iteration of the two formulas (we apply it to $J_\lambda\varphi$, $J_\lambda^2\varphi$, etc.), we obtain, for $k \geq 1$,

$$J_\lambda^k \varphi = \mathbb{E}^\sharp P_{\sigma_k}\varphi, \quad J_\lambda^k\varphi(X_t) = \mathbb{E}^\sharp\mathbb{E}\left[\varphi(X_{\sigma_k+t}) \middle| \mathcal{F}_t\right], \tag{5.64}$$

where $\sigma_k = \tau_1 + \cdots + \tau_k$ for $\tau_1, \ldots, \tau_k$ some i.i.d. $\mathcal{E}(\lambda)$ independent random variables. from $(\mathcal{F}_t)_{t\geq 0}$. Take now $\lambda = N$, where $N \to +\infty$ and $k = [Ns]$ for a given $s > 0$. By the weak law of large numbers, we have $\sigma_k \to s$ in probability (for $\mathbb{P}^\sharp$). Consequently, the limit $[N \to +\infty]$ of the first equality in (5.64) gives

$$J_N^{[ns]}\varphi \xrightarrow{\pi} P_s\varphi. \tag{5.65}$$

The map $\theta\colon \sigma \mapsto \mathbb{E}[\varphi(X_{\sigma+t})|\mathcal{F}_t]$ is continuous since, for $\sigma' \geq \sigma$, and by the martingale property,

$$\left|\theta(\sigma') - \theta(\sigma)\right| = \left|\int_{t+\sigma}^{t+\sigma'} \mathbb{E}[\mathcal{L}\varphi(X_s)|\mathcal{F}_t]ds\right| \leq \|\mathcal{L}\varphi\|_{\mathrm{BM}(E)}(\sigma' - \sigma).$$

Consequently, at the limit $[N \to +\infty]$ in the second identity in (5.64), we get (5.49). We have supposed $\varphi \in \mathcal{D}(\mathcal{L})$, but $\mathcal{D}(\mathcal{L})$ is π-dense in $C_b(E)$ (*cf.* Proposition 4.6), therefore (5.49) holds true when φ is an arbitrary element of $C_b(E)$. □

6 Diffusion approximation in finite dimension

Let $d, k \geq 0$, let $f \in C_b^1(\mathbb{R}^d; \mathbb{R}^d)$ and $g \in C_b^2(\mathbb{R}^k \times \mathbb{R}^d; \mathbb{R}^d)$, where C_b^k denote the set of functions with continuous and bounded derivatives for all orders from 0 to k. Let $(\mathcal{F}_t)_{t\geq 0}$ be a complete filtration and let $(m_t(n))$ be a collection of càdlàg Markov process on $\mathbb{R}^k$ such that, for every $\mathcal{F}_0$-measurable random variable $n\colon \Omega \to \mathbb{R}^k$,

$$\mathbb{P}(m_0(n) = n) = 1, \tag{6.1}$$
$$\mathbb{E}\left[\varphi(m_{t+s}(n))|\mathcal{F}_t\right] = \mathbb{E}\left[(P_t\varphi)(m_t(n))\right], \tag{6.2}$$

for all $\varphi \in \mathrm{BM}(\mathbb{R}^k)$, for all $0 \leq s, t$, where $(P_t)_{t\geq 0}$ is a Markov semi-group associated to a transition function $\{Q_t; t \geq 0\}$, with $(P_t)_{t\geq 0}$ satisfying (4.16). We assume that, for every $t \geq 0$, the map

$$\Omega \times \mathbb{R}^k \to \mathbb{R}^k, \quad (\omega, n) \mapsto m_t(n), \tag{6.3}$$

is measurable. We assume also that $(m_t(n))$ has the invariant measure $\nu \in \mathcal{P}(\mathbb{R}^k)$. More precisely, we assume that there is a $\mathcal{F}_0$-measurable random variable $\bar{n}$ having the law ν such that $\bar{m}_t := m_t(\bar{n})$ is a stationary process (note that, in virtue of Theorem 4.2, this amounts to require $\mathrm{Law}(\bar{m}_t) = \nu$ for all $t \geq 0$). Let $x \in \mathbb{R}^d$. Our aim in this section is to find the limit when $\varepsilon \to 0$ of the solution X_t^ε to the Cauchy problem

$$\frac{dX_t^\varepsilon}{dt} = f(X_t^\varepsilon) + \frac{1}{\varepsilon}g(\bar{m}_t^\varepsilon, X_t^\varepsilon), \tag{6.4}$$
$$X_0^\varepsilon = x. \tag{6.5}$$

In (6.4) the process $\bar{m}_t^\varepsilon$ is the rescaled process

$$\bar{m}_t^\varepsilon = \bar{m}_{\varepsilon^{-2}t}. \tag{6.6}$$

The plan of this section is the following one. In Section 6.1, we explain what is the framework (hypotheses on $(m_t(n))$) in which (6.4) may have a limit. In Section 6.1

we analyse the Markov property for $(X^\varepsilon_t, m^\varepsilon_t(n))$ and give the associated generator $\mathcal{L}^\varepsilon$. In Section 6.3, we find the limit generator $\mathcal{L}$ by a perturbed test function method. In Section 6.4, we prove the tightness of (X^ε_t). In Section 6.5, we display a limit martingale problem. In Section 6.6, we identity the limit problem as a stochastic differential equation. This uses the theory of stochastic differential equation, whose treatment is reported to Section 8. The main result of this section is Theorem 6.13.

6.1 Mixing hypothesis

6.1.1 Hypotheses on the driving stochastic process

Assume that $(\bar{m}_t)$ has the following ergodic property: for all $n \in \mathbb{R}^k$, for all $\psi \in \mathrm{BM}(\mathbb{R}^k)$,

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T \mathbb{E}\psi(m_t(n))dt \to \langle \nu, \psi\rangle. \tag{6.7}$$

Taking $\psi = g(\cdot, x)$, and $T = \varepsilon^{-2}t$, we see that

$$\frac{1}{\varepsilon}\int_0^t \mathbb{E}[g(m^\varepsilon_t(n), x)]dt \simeq \frac{t}{\varepsilon}\langle \nu, g(\cdot, x)\rangle.$$

The solution X^ε_t to (6.4) will be singular when $\varepsilon \to 0$, unless the first moment vanishes:

$$\int_{\mathbb{R}^k} g(n,x)d\nu(n) = \mathbb{E}[g(\bar{m}_t, x)] = 0, \quad \forall x \in \mathbb{R}^d. \tag{6.8}$$

Our general framework will be the following one: we will assume that the condition (6.8) is satisfied and that $(m_t(n))_{t\geq 0}$ has the following *mixing* property: there exists a non-negative non-increasing function $\gamma_{\mathrm{mix}} \in L^1(\mathbb{R}_+)$ such that, for all n, n' $\mathcal{F}_0$-measurable random variables, there is a *coupling* $(m^*_t(n), m^*_t(n'))_{t\geq 0}$ of $(m_t(n), m_t(n'))_{t\geq 0}$ such that

$$\mathbb{E}\left[|m^*_t(n) - m^*_t(n')|\right] \leq \gamma_{\mathrm{mix}}(t)\mathbb{E}[|n-n'|]. \tag{6.9}$$

We have used the

Definition 6.1 (coupling)**.** Let E be a separable Banach space and $(X_t)_{t\geq 0}$, $(Y_t)_{t\geq 0}$ two E-valued stochastic processes. The couple $(\tilde{X}_t, \tilde{Y}_t)_{t\geq 0}$ is a *coupling* of $(X_t, Y_t)_{t\geq 0}$ if $(\tilde{X}_t)_{t\geq 0}$ and $(\tilde{Y}_t)_{t\geq 0}$ have the same law as $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ respectively.

If $(\tilde{X}_t, \tilde{Y}_t)$ is a coupling of (X_t, Y_t), their joint law may be different, and that is the potential interest of a coupling. The estimate (6.9) can be expressed in terms of the joint law of $(m^*_t(n))$ and $(m^*_t(n'))$. See the examples in Section 6.1.3. See also Appendix A on the problem of maximal coupling.
Assume

$$M_1 = \mathbb{E}|\bar{n}| = \mathbb{E}|\bar{m}_t| = \int_{\mathbb{R}^k} |n|d\nu(n) < +\infty. \tag{6.10}$$

If $n \in \mathbb{R}^k$ and $\varphi \in \mathrm{BM}(\mathbb{R}^k)$ is Lipschitz continuous, (6.9) gives

$$|P_t\varphi(n) - \langle\varphi, \nu\rangle| = |\mathbb{E}[\varphi(m^*_t(n)) - \varphi(m^*_t(\bar{n}))]| \leq \gamma_{\mathrm{mix}}(t)\mathrm{Lip}(\varphi)(|n| + M_1), \tag{6.11}$$

and, for $n' \in \mathbb{R}^k$,

$$|P_t\varphi(n) - P_t\varphi(n')| = |\mathbb{E}[\varphi(m_t^*(n)) - \varphi(m_t^*(n'))]| \leq \gamma_{\text{mix}}(t)\text{Lip}(\varphi)|n - n'|. \tag{6.12}$$

We will make a much stronger hypothesis than (6.10). We assume that there exists $\kappa > 0$ such that the closed ball $\bar{B}_\kappa$ of center 0 and radius κ is stable by evolution along the process: for any $\mathcal{F}_0$-measurable random variable n, we have

$$n \in \bar{B}_\kappa \text{ a.s.} \Rightarrow m_t(n) \in \bar{B}_\kappa \text{ a.s. for all } t \geq 0. \tag{6.13}$$

We deduce from (6.9) and (6.13) that $\bar{m}_t \in \bar{B}_\kappa$ a.e. Indeed, if φ is Lipschitz continuous, non-negative and vanishes on $\bar{B}_\kappa$ and n is given in $\bar{B}_\kappa$, then, due to (6.11),

$$\mathbb{E}\varphi(\bar{m}_t) = \langle \varphi, \nu \rangle = \lim_{s \to +\infty} P_s\varphi(n) = 0,$$

since $P_s\varphi(n) = 0$ for all $s \geq 0$. Consequently, we have

$$|\bar{m}_t| \leq \kappa \text{ a.s. for all } t \geq 0. \tag{6.14}$$

We will now take $\bar{B}_\kappa$ as a state space for the process $(m_t(n))$.

6.1.2 The Poisson equation

Later (see Section 6.3), we will need to solve the Poisson equation associated to (m_t), Equation (6.15) below.

Proposition 6.1 (The Poisson equation)**.** *Let $\mathcal{A}$ denote the generator of (m_t). Let $\varphi \in \text{Lip}(\bar{B}_\kappa)$ satisfy the cancellation condition $\langle \nu, \varphi \rangle = 0$. Then the Poisson equation*

$$-\mathcal{A}\psi = \varphi \tag{6.15}$$

has a unique solution $\psi \in \mathcal{D}(\mathcal{A})$ such that $\langle \nu, \psi \rangle = 0$, given by

$$\psi(n) = R_0\varphi(n) = \int_0^\infty P_t\varphi(n)dt. \tag{6.16}$$

We have also the bound

$$\text{Lip}(\psi) \leq \|\gamma_{\text{mix}}\|_{L^1(\mathbb{R}_+)}\text{Lip}(\varphi), \tag{6.17}$$

for ψ given by (6.16).

Proof of Proposition 6.1. Since $\langle \varphi, \nu \rangle = 0$, the estimate (6.11) shows that the integral defining ψ in (6.16) is convergent. The estimate (6.17) follows from (6.12). It is also simple to show that $\psi \in \mathcal{D}(\mathcal{A})$, and that ψ satisfies (6.15). There remains to show uniqueness. We start from the identity (it follows from (4.4))

$$P_t\psi(n) = \psi(n) + \int_0^t P_s\mathcal{A}\psi(n)ds. \tag{6.18}$$

If $\mathcal{A}\psi = 0$ and $\langle \psi, \nu \rangle = 0$, we deduce from (6.11) that

$$|\psi(n)| \leq \gamma_{\text{mix}}(t)\text{Lip}(\varphi)(|n| + M_1).$$

At the limit $t \to +\infty$, we obtain $\psi(n) = 0$. $\square$

Remark 6.1. If $\psi \in \mathcal{D}(\mathcal{A})$ satisfies $\mathcal{A}\psi = 0$, then ψ is constant. This follows from the uniqueness part of Proposition 6.1 applied to $\psi - \langle\psi, \nu\rangle$.

6.1.3 Some examples

We give some classical examples of processes $(m_t(n))$ satisfying (6.8), (6.9), (6.10). Our first example is the Ornstein-Uhlenbeck process. Maybe the more classical definition of the Ornstein-Uhlenbeck process is that it should solve the stochastic differential equation

$$dm_t(n) = -m_t(n)dt + \sqrt{2}dB_t, \quad m_0(n) = n, \tag{6.19}$$

where (B_t) is a k-dimensional Wiener process. Since we have not seen stochastic differential equations yet, we take the following definition:

$$m_t(n) = e^{-t}n + B_{r(t)}, \quad r(t) = 1 - e^{-2t}. \tag{6.20}$$

Indeed, the solution to (6.19) has the same law as the right-hand side of (6.20). The invariant measure is the $\mathcal{N}(0,1)$. The cancellation condition (6.8) is then satisfied if $g(\cdot, x)$ is odd, for example. A coupling of $m_t(n)$ and $m_t(n')$ that gives (6.9) is the synchronous coupling, that uses the same realization of Wiener process for both trajectories. Here this amounts to no coupling at all: we have

$$\mathbb{E}|m_t(n) - m_t(n')| = e^{-t}\mathbb{E}|n - n'|,$$

hence (6.9) with $\gamma_{\mathrm{mix}}(t) = e^{-t}$. It is clear that (6.10) is satisfied. Note also that

$$\mathcal{A}\varphi(n) = -n \cdot \nabla_n\varphi(n) + \Delta_n\varphi(n).$$

However, (6.13) is *not* satisfied since a Gaussian is not compactly supported. Our second example is a Markov jump process. Let ν be a given probability measure on $\mathbb{R}^k$ supported in $\bar{B}_\kappa$ such that

$$\int_{\mathbb{R}^k} n d\nu(n) = 0.$$

Let $(\mathcal{F}^0_k)_{k\in\mathbb{N}}$ be a filtration indexed by $\mathbb{N}$. Let $(m^0_k(n))_{k\geq 0}$ be a $(\mathcal{F}^0_k)$-Markov chain on $\mathbb{R}^k$ having the invariant measure ν and satisfying the following mixing property: there exists $\gamma \in (0,1)$ and $C \geq 0$ such that, for all n, n' $\mathcal{F}^0_0$-measurable, there is a coupling $(m^{0,*}_k(n), m^{0,*}_k(n'))$ of $(m^0_k(n), m^0_k(n'))$ such that

$$\mathbb{E}|m^{0,*}_k(n) - m^{0,*}_k(n')| \leq C\gamma^k\mathbb{E}|n - n'|. \tag{6.21}$$

Assume also that $\bar{B}_\kappa$ is stable by $n \mapsto m^0_k(n)$ for every k. Let $(T_j)_{j\in\mathbb{N}}$ be a Poisson process of constant rate 1 independent on $(\mathcal{F}^0_k)_{k\geq 1}$. For $t \geq 0$ and n a $\mathcal{F}^0_0$-measurable function, let $S_k = T_1 + \cdots + T_k$ and let

$$m_t(n) = n\mathbf{1}_{t<T_1} + \sum_{k\geq 1} m^0_k(n)\mathbf{1}_{S_k\leq t<S_{k+1}}. \tag{6.22}$$

By Exercise 4.11, $(m_t(n))$ is a Markov process for the filtration $\mathcal{F}_t$ generated by $\mathcal{F}_t^m$ and the natural filtration of the Poisson process. Note that $\mathcal{F}_0 = \mathcal{F}_0^0$. To obtain (6.9), we use the coupling

$$m_t^*(n) = n\mathbf{1}_{t<T_1} + \sum_{k\geq 1} m_k^{0,*}(n)\mathbf{1}_{S_k\leq t<S_{k+1}}. \tag{6.23}$$

This is again a synchronous coupling since the time of the jumps are the same for both trajectories. By independence and by (6.21), we have

$$\mathbb{E}|m_t^*(n) - m_t^*(n')| \leq \mathbb{P}(t<T_1)\mathbb{E}|n-n'| + \sum_{k\geq 1} C\gamma^k\mathbb{E}|n-n'|\mathbb{P}(S_k\leq t<S_{k+1}).$$

Since $\mathbb{P}(t<T_1) = e^{-t}$, $\mathbb{P}(S_k\leq t<S_{k+1}) = e^{-t}$, we obtain (6.9) with

$$\gamma_{\mathrm{mix}}(t) = \left[1+\frac{C\gamma}{1-\gamma}\right]e^{-t}.$$

It is clear that (6.13) is satisfied here.

6.2 Markov property

6.2.1 Resolution of the ODE

We are interested in the resolution of (6.4)-(6.5). Since we work at fixed $\varepsilon>0$ for the moment. We will first consider the Cauchy Problem

$$\frac{dX_t}{dt} = f(X_t) + g(q_t, X_t), \tag{6.24}$$

$$X_0 = x, \tag{6.25}$$

where (q_t) is a given càdlàg function.

Proposition 6.2. *Let $T>0$. On the interval $[0,T]$, the problem (6.24)-(6.25) has a unique solution $X\in C([0,T];\mathbb{R}^d)$. If $\tilde{q}_t$ is another càdlàg function and $\tilde{X}\in C([0,T];\mathbb{R}^d)$ the associated solution to (6.24)-(6.25), then*

$$|X_t-\tilde{X}_t| \leq \mathrm{Lip}(g)e^{(\mathrm{Lip}(f)+\mathrm{Lip}(g))T}\int_0^t |q_s-\tilde{q}_s|ds. \tag{6.26}$$

Proof of Proposition 6.2. The existence and uniqueness of $X\in C([0,T];\mathbb{R}^d)$ solution to (6.24)-(6.25) follows from the Cauchy-Lipschitz theorem. To obtain (6.26), we write

$$\begin{aligned}|X_t-\tilde{X}_t| &\leq \int_0^t \left[|f(X_s)-f(\tilde{X}_s)| + |g(q_s,X_s)-g(\tilde{q}_s,\tilde{X}_s)|\right]ds\\ &\leq (\mathrm{Lip}(f)+\mathrm{Lip}(g))\int_0^t|X_s-\tilde{X}_s|ds + \mathrm{Lip}(g)\int_0^t|q_s-\tilde{q}_s|ds.\end{aligned}$$

The Gronwall Lemma gives

$$|X_t-\tilde{X}_t| \leq \mathrm{Lip}(g)\int_0^t e^{(\mathrm{Lip}(f)+\mathrm{Lip}(g))(t-s)}|q_s-\tilde{q}_s|ds,$$

and (6.26) follows. □

6.2.2 Markov property

Let us denote by $X(t,0;x,(q_\sigma)_{\sigma\in[0,t]})$ the solution to (6.24)-(6.25). More generally, we can denote by $X(t,s;x,(q_\sigma)_{\sigma\in[s,t]})$ the value at time t of the solution to (6.24) on $[s,t]$ starting from x at time s. By uniqueness, we have the semi-group property

$$X(t,s;x,(q_\sigma)_{\sigma\in[s,t]})=X(t,\tau;y,(q_\sigma)_{\sigma\in[\tau,t]}),\quad y=X(\tau,s;x,(q_\sigma)_{\sigma\in[s,\tau]}),\tag{6.27}$$

for all $s\leq\tau\leq t$. Assume now that $(q_t(z))$ is a càdlàg Markov process relatively to a complete filtration $(\mathcal{G}_t)$. Then we have the following result.

Proposition 6.3 (Markov property)**.** *The process (X_t,q_t) is a Markov process, relatively to $(\mathcal{G}_t)$, with transition operator (Π_t) given by*

$$\Pi_t\varphi(x,z)=\mathbb{E}\varphi\left[X(t,0;x,(q_s(z))_{s\in[0,t]}),q_t(z)\right].\tag{6.28}$$

To prove Proposition 6.3, we will need the following lemma.

Lemma 6.4. *Assume that, for all $t\geq 0$,*

$$\Omega\times\mathbb{R}^k\to\mathbb{R}^k,\quad(\omega,z)\mapsto q_t^\omega(z),\tag{6.29}$$

is measurable. Let Π_t be defined by (6.28). Then $\Pi_t\colon \mathrm{BM}(\mathbb{R}^d\times\mathbb{R}^k)\to\mathrm{BM}(\mathbb{R}^d\times\mathbb{R}^k)$ and (Π_t) satisfies the points 1, 2, 3 of Definition 4.5.

Note: we do not assert at that point that (Π_t) is a π-contraction semi-group since the semi-group property will be established later, in the proof of Proposition 6.3.

Proof of Lemma 6.4. Note first that (6.29) implies that

$$\Omega\times\mathbb{R}^k\to L^1((0,t);\mathbb{R}^k),\quad(\omega,z)\mapsto(q_s^\omega(z))_{s\in[0,t]},\tag{6.30}$$

is measurable. Indeed, for $n\in\mathbb{N}^*$, set $s_k=\frac{kt}{n}$ and define

$$q_s^{n,\omega}(z)\sum_{k=0}^{n-1}q_{s_{k+1}}^\omega(z)\mathbf{1}_{[s_k,s_{k+1})}(s)$$

We have $q_s^{n,\omega}(z)\to q_{s+}^\omega(z)=q_s^\omega(z)$ for every $s\in[0,t]$ since $s\mapsto q_s^\omega(z)$ is càdlàg, and thus

$$(q_s^{n,\omega}(z))_{s\in[0,t]}\to(q_s^\omega(z))_{s\in[0,t]}$$

in $L^1(0,t)$. We also have

$$|X_t(x)-x|\leq(\|f\|_{\mathrm{BM}(\mathbb{R}^d)}+\|g\|_{\mathrm{BM}(\mathbb{R}^k\times\mathbb{R}^d)})t.\tag{6.31}$$

Together with (6.26), this shows that

$$(\omega,x,z)\mapsto(X(t,0;x,(q_s(z))_{s\in[0,t]}),q_t(z))$$

is measurable. It follows that $\Pi_t\colon \mathrm{BM}(\mathbb{R}^d\times\mathbb{R}^k)\to\mathrm{BM}(\mathbb{R}^d\times\mathbb{R}^k)$. We have

$$\Pi_t\varphi(x,z)=\langle\varphi,\lambda_t\rangle,\quad \lambda_t:=\mathrm{Law}(X(t,0;x,(q_s(z))_{s\in[0,t]}),q_t(z)). \tag{6.32}$$

From (6.32) we deduce the points 2 and 3 of Definition 4.5. To prove the point 1 (stochastic continuity), we simply use the dominated convergence theorem. Indeed, almost-surely,

$$(X(t,0;x,(q_s(z))_{s\in[0,t]}),q_t(z))\to(x,z)$$

by (6.31) and the càdlàg property of $(q_t(z))$. □

Proof of Proposition 6.3. First, it is clear that (X_t) is $(\mathcal{G}_t)$-adapted. A way to see this is to write $(X_t)_{t\in[0,T]}$ as the solution to a fixed point equation $X=\mathcal{T}(X)$ on $C([0,T],\mathbb{R}^d)$, where, for an adequate[7] norm on $C([0,T],\mathbb{R}^d)$, $\mathcal{T}$ is a contraction. It follows that $X=\lim_{n\to+\infty}X^n$ a.s. in $C([0,T],\mathbb{R}^d)$, where X^n is the sequence defined by the iteration $X^{n+1}=\mathcal{T}(X^n)$, $X^0_t=x$. Since each X^n_t is $\mathcal{G}_t$-measurable, so is X_t (Proposition 4.7). By (6.27), we have

$$\begin{aligned}\varphi&\left[X(t+s,0;x,(q_\sigma(z))_{\sigma\in[0,t+s]}),q_{t+s}(z)\right]\\&=\varphi\left[X(t+s,t;y,(q_\sigma(z))_{\sigma\in[t,t+s]}),q_{t+s}(z)\right],\end{aligned} \tag{6.33}$$

where $y=X(t,0;x,(q_\sigma(z))_{\sigma\in[0,t]})$. Taking the conditional expectation of (6.33), we see that we will obtain the Markov property

$$\mathbb{E}\left[\varphi(X_{t+s},q_{t+s}(z))|\mathcal{G}_t\right]=(\Pi_s)(X_t,q_t(z)),\quad X_t=X(t,0;x,(q_\sigma)_{\sigma\in[0,t]}), \tag{6.34}$$

if we establish that

$$\mathbb{E}\left[\varphi\left[X(t+s,t;Y,(q_\sigma(z))_{\sigma\in[t,t+s]}),q_{t+s}(z)\right]\mid\mathcal{G}_t\right]=(\Pi_s\varphi)(Y,q_t(z)) \tag{6.35}$$

for all $\mathcal{G}_t$-measurable random variable Y. We consider first a continuous and bounded function φ. Let $(s_i)_{0,N}$ be a regular subdivision of the interval $[0,s]$ and let

$$\tilde q^N_\sigma(z)=\sum_{i=0}^{N-1}q_{s_i+t}(z)\mathbf{1}_{[s_i+t,s_{i+1}+t)}(\sigma),\quad \sigma\in[t,s+t].$$

We claim that it is sufficient to consider (6.35) with the path $(q_\sigma(z))_{\sigma\in[t,t+s]})$ replaced by the path $(\tilde q^N_\sigma(z))_{\sigma\in[t,t+s]})$. Indeed, $(\tilde q^N_\sigma(z))$ is converging to $(q_\sigma(z))$ when $N\to+\infty$ in $L^1(t,s+t)$. Indeed, we have $\tilde q^N_\sigma(z)=q_{s_i(\sigma)+t}(z)$ where $s_i(\sigma)+t\leq\sigma<s_{i+1}(\sigma)+t$. Since $s_{i+1}(\sigma)=s_i(\sigma)+\frac{s}{N}$, we see that $s_i(\sigma)+t$ is converging to σ from below when $N\to+\infty$ and, consequently, $\tilde q^N_\sigma(z)$ is converging to $q_{\sigma-}(z)$ when $N\to+\infty$, for all $\sigma\in[t,s+t]$. By $q_{\sigma-}(z)$ we denote the limit from the left of the càdlàg function $\sigma\mapsto q_\sigma(z)$. It coincides with the value $q_\sigma(z)$, except at a countable set of points σ. The L^1-convergence follows

[7] take $\|X\|=\sup_{t\in[0,T]}e^{-(\mathrm{Lip}(f)+\mathrm{Lip}(g))t}|X_t|$

by the dominated convergence theorem. Those arguments show that the left-hand side of (6.35) is the limit when $N \to +\infty$ of

$$\mathbb{E}\left[\varphi\left[X(t+s,t;Y,(\tilde{q}^N_\sigma(z))_{\sigma\in[t,t+s]}),q_{t+s}(z)\right] \mid \mathcal{G}_t\right]. \tag{6.36}$$

By Proposition 6.2, there is a continuous function ξ such that

$$X(t+s,t;Y,(\tilde{q}^N_\sigma(z))_{\sigma\in[t,t+s]}) = \xi(s;Y,q_{t+s_0}(z),\dots,q_{t+s_{N-1}}(z)), \tag{6.37}$$

$$\xi(s;Y,q_{s_0}(z),\dots,q_{s_{N-1}}(z)) = X(s,0;Y,(\hat{q}^N_\sigma(z))_{\sigma\in[0,s]}), \tag{6.38}$$

where

$$\hat{q}^N_\sigma(z) = \sum_{i=0}^{N-1} q_{s_i}(z)\mathbf{1}_{[s_i,s_{i+1})}(\sigma), \quad \sigma\in[0,s].$$

Using (6.37)-(6.38) and the Markov property for (q_σ), we obtain

$$\begin{aligned}(6.36) &= \mathbb{E}\left[\varphi\left[\xi(s;Y,q_{t+s_0}(z),\dots,q_{t+s_{N-1}}(z)),q_{t+s}(z)\right] \mid \mathcal{G}_t\right] = \Psi_s(Y,q_t(z)),\\ \Psi_s(x,z) &:= \mathbb{E}\varphi\left[\xi(s;x,q_{s_0}(z),\dots,q_{s_{N-1}}(z)),q_s(z)\right]\\ &= \mathbb{E}\left[\varphi\left[X(s,0;x,(\hat{q}^N_\sigma(z))_{\sigma\in[0,s]}),q_s(z)\right]\right].\end{aligned}$$

By taking the limit $[N\to+\infty]$, we obtain (6.35), and hence (6.34), with the restriction that φ is continuous. We deduce that, for $\varphi\in C_b(\mathbb{R}^d\times\mathbb{R}^k)$ and $s,t\geq 0$,

$$\begin{aligned}\Pi_{t+s}\varphi(x,z) &= \mathbb{E}\left[\varphi(X_{t+s}(x),q_{t+s}(z))\right] = \mathbb{E}\left[\mathbb{E}\left[\varphi(X_{t+s}(x),q_{t+s}(z))|\mathcal{G}_t\right]\right]\\ &= \mathbb{E}\left[(\Pi_s\varphi)(X_t(x),q_t(z))\right] = (\Pi_t\circ\Pi_s\varphi)(x,z).\end{aligned} \tag{6.39}$$

Consider the two maps

$$\varphi\mapsto\Pi_{t+s}\varphi(x,z), \quad \varphi\mapsto(\Pi_t\circ\Pi_s\varphi)(x,z).$$

They preserve the positivity and fix the constants and are continuous for π-convergence due to Lemma 6.4. It follows (see the proof or Proposition 4.3) that

$$A\mapsto\Pi_{t+s}\mathbf{1}_A(x,z), \quad A\mapsto(\Pi_t\circ\Pi_s\mathbf{1}_A)(x,z)$$

are both probability measures on $\mathcal{B}(\mathbb{R}^d\times\mathbb{R}^k)$. These two measures coincide when tested against functions of $C_b(\mathbb{R}^d\times\mathbb{R}^k)$ and $C_b(\mathbb{R}^d\times\mathbb{R}^k)$ is a separating class (Proposition 2.10). These two measures are therefore equal, and we deduce that (6.39) is satisfied for any function $\varphi\in\mathrm{BM}(\mathbb{R}^d\times\mathbb{R}^k)$. This gives the semi-group property $\Pi_{t+s}=\Pi_t\circ\Pi_s$. By Lemma 6.4 and Proposition 4.3, we deduce that (Π_t) is a Markov semi-group. To conclude, we have to show that (6.34) is satisfied not only for continuous functions, but for all $\varphi\in\mathrm{BM}(\mathbb{R}^d\times\mathbb{R}^k)$. Our aim is to prove that, given B a $\mathcal{G}_t$-measurable set,

$$\mathbb{E}\left[\mathbf{1}_B\varphi(X_{t+s},q_{t+s}(z))\right] = \mathbb{E}\left[\mathbf{1}_B(\Pi_s)(X_t,q_t(z))\right]. \tag{6.40}$$

We do the same reasoning as above. Both members of (6.40) define some Radon measures (of total mass $\mathbb{P}(B)$) that coincide when tested against continuous bounded functions. An argument of separating class gives the conclusion. □

We have worked on the system (6.24)-(6.25). Let $\mathcal{F}_t^\varepsilon = \mathcal{F}_{\varepsilon^{-2}t}$. Replacing g by $\frac{1}{\varepsilon}g$ and taking m_t^ε as a driving process, we obtain the following result for the solution

$$X_t^\varepsilon = X^\varepsilon(t, 0; x, (m_\sigma^\varepsilon(n))_{\sigma\in[0,t]})$$

to

$$\frac{dX_t^\varepsilon}{dt} = f(X_t^\varepsilon) + \frac{1}{\varepsilon}g(m_t^\varepsilon(n), X_t^\varepsilon), \quad X_0^\varepsilon = x. \tag{6.41}$$

The solution $(X_t^\varepsilon, m_t^\varepsilon)$ is a Markov process relatively to $(\mathcal{F}_t^\varepsilon)_{t\geq 0}$ with transition operator Π_t^ε given by

$$\Pi_t^\varepsilon\varphi(x, n) = \mathbb{E}\varphi\left[X^\varepsilon(t, 0; x, (m_\sigma^\varepsilon(n))_{\sigma\in[0,t]}), m_t^\varepsilon(n)\right], \tag{6.42}$$

for $\varphi \in \mathrm{BM}(\mathbb{R}^d \times \bar{B}_\kappa)$, $(x, n) \in \mathbb{R}^d \times \bar{B}_\kappa$.

6.2.3 Generator

Definition 6.2 (Admissible test-function)**.** A continuous bounded function $\varphi \colon \mathbb{R}^d \times \bar{B}_\kappa \to \mathbb{R}$ is said to be an *admissible test-function* if

1. for all $x \in \mathbb{R}^d$, $\varphi(x, \cdot) \in \mathcal{D}(\mathcal{A})$ and $(x, n) \mapsto \mathcal{A}\varphi(x, n)$ is bounded,
2. for all $n \in \bar{B}_\kappa$, $\varphi(\cdot, n)$ is of class C^1 on $\mathbb{R}^d$ and $(x, n) \mapsto \nabla_x\varphi(x, n)$ is bounded,
3. for all $x \in \mathbb{R}^d$, $\nabla_x\varphi(x, \cdot) \in \mathcal{D}(\mathcal{A})$, and $(x, n) \mapsto \mathcal{A}\nabla_x\varphi(x, n)$ is bounded.

Remark 6.2 (Poisson's equation with parameter). It is clear from the proof of Proposition 6.1 that, if φ is a Lipschitz continuous admissible test-function such that

$$\langle\varphi(x, \cdot), \nu\rangle = 0$$

for all x, then

$$\psi(n, x) = R_0\varphi(n, x) = \int_0^\infty P_t\varphi(n, x)dt,$$

is also an admissible test-function and satisfies the Poisson equations $-\mathcal{A}\psi(\cdot, x) = \varphi(\cdot, x)$ and $-\mathcal{A}\nabla_x\psi(\cdot, x) = \nabla_x\varphi(\cdot, x)$ for all $x \in \mathbb{R}^d$.

Proposition 6.5 (Generator)**.** *Let $\mathcal{L}^\varepsilon$ be the generator associated to the transition semigroup $(\Pi_t^\varepsilon)_{t\geq 0}$ given by (6.42). If φ is an admissible test-function, then φ is in the domain of $\mathcal{L}^\varepsilon$ and*

$$\mathcal{L}^\varepsilon\varphi(x, n) = \frac{1}{\varepsilon^2}\mathcal{A}\varphi(x, n) + \frac{1}{\varepsilon}g(n, x)\cdot\nabla_x\varphi(x, n) + f(x)\cdot\nabla_x\varphi(x, n), \tag{6.43}$$

for all $(x, n) \in \mathbb{R}^d \times \bar{B}_\kappa$.

Proof of Proposition 6.5. Let φ be an admissible test-function. Note first that $\mathcal{L}^\varepsilon\varphi$ as defined by (6.43) is a bounded measurable function. We want to show that

$$\Pi_t^\varepsilon\varphi = \varphi + t\mathcal{L}^\varepsilon\varphi + \eta^\varepsilon(t)t, \tag{6.44}$$

where (η^ε) is π-converging to 0 on $\mathbb{R}^d \times \bar{B}_\kappa$. We split the difference $\Pi_t^\varepsilon\varphi - \varphi$ into the sum of the two terms

$$\mathbb{E}\varphi\left(X_t^\varepsilon, m_t^\varepsilon(n)\right) - \mathbb{E}\varphi(x, m_t^\varepsilon(n)), \tag{6.45}$$

and $\mathbb{E}\varphi(x, m_t^\varepsilon(n)) - \varphi(x,n)$. For this last term, we have

$$\mathbb{E}\varphi(x, m_t^\varepsilon(n)) - \varphi(x,n) = P_t^\varepsilon\varphi(x,n) - \varphi(x,n) = \frac{t}{\varepsilon^2}\mathcal{A}\varphi(x,n) + t\zeta^\varepsilon(t), \tag{6.46}$$

where

$$\zeta^\varepsilon(t)\colon (x,n) \mapsto \frac{1}{t}\int_0^{t/\varepsilon^2} \left[P_s\mathcal{A}\varphi(x,n) - \mathcal{A}\varphi(x,n)\right] ds = \int_0^{1/\varepsilon^2} \left[P_{st}\mathcal{A}\varphi(x,n) - \mathcal{A}\varphi(x,n)\right] ds$$

is π-converging to 0 since, by (4.16), $P_s\mathcal{A}\varphi(x,n) \to \mathcal{A}\varphi(x,n)$ when $s \to 0$. The first term (6.45) is

$$\mathbb{E}\int_0^t \nabla_x\varphi\left(X_s^\varepsilon, m_t^\varepsilon(n)\right)\cdot \dot{X}_s^\varepsilon ds = \mathbb{E}\int_0^t \nabla_x\varphi\left(X_s^\varepsilon, m_t^\varepsilon(n)\right)\cdot H^\varepsilon(X_s^\varepsilon, m_s^\varepsilon) ds, \tag{6.47}$$

where $H^\varepsilon(x,n) = f(x) + \frac{1}{\varepsilon}g(n,x)$. To obtain the asymptotic expansion of (6.45), we introduce the partial maps

$$\psi(x,n;n') = \nabla_x\varphi(x,n')\cdot H^\varepsilon(x,n), \quad \theta(n;x'n') = \nabla_x\varphi(x',n)\cdot H^\varepsilon(x',n').$$

We have then (6.47) $= t\nabla_x\varphi(x,n)\cdot H^\varepsilon(x,n) + t\xi_1^\varepsilon(t) + t\xi_2^\varepsilon(t)$, with

$$\xi_1^\varepsilon(t) = \frac{1}{t}\int_0^t \mathbb{E}\left[\nabla_x\varphi\left(X_s^\varepsilon, m_t^\varepsilon(n)\right)\cdot H^\varepsilon(X_s^\varepsilon, m_s^\varepsilon) - \nabla_x\varphi\left(X_s^\varepsilon, n\right)\cdot H^\varepsilon(X_s^\varepsilon, m_s^\varepsilon)\right] ds,$$

which we rewrite as

$$\begin{aligned}\xi_1^\varepsilon(t) &= \frac{1}{t}\int_0^t \left[P_{\varepsilon^{-2}t}\theta(n;x',n') - \theta(n;x',n')\right]|_{(x',n')=(X_s^\varepsilon, m_s^\varepsilon(n))} ds \\ &= \int_0^1 \left[P_{\varepsilon^{-2}t}\theta(n;x',n') - \theta(n;x',n')\right]|_{(x',n')=(X_{st}^\varepsilon, m_{st}^\varepsilon(n))} ds,\end{aligned}$$

and

$$\xi_2^\varepsilon(t) = \frac{1}{t}\int_0^t \mathbb{E}\left[\nabla_x\varphi\left(X_s^\varepsilon, n\right)\cdot H^\varepsilon(X_s^\varepsilon, m_s^\varepsilon) - \nabla_x\varphi\left(x,n\right)\cdot H^\varepsilon(x,n)\right] ds,$$

equal to

$$\xi_2^\varepsilon(t) = \int_0^1 \left[\Pi_{st}^\varepsilon\psi(x,n;n') - \psi(x,n;n')\right]|_{n'=n} ds.$$

By dominated convergence, $\xi_2^\varepsilon(t) \to 0$ when $t \to 0$ and is a bounded function of (x,n). To get a similar result for $\xi_1^\varepsilon(t)$, we need a convergence $P_t\theta(n;x',n') \to \theta(n;x',n')$ that is uniform in (x',n'). This is the case since $\nabla_x\varphi(x,\cdot) \in \mathcal{D}_1(\mathcal{A})$, and thus

$$|P_t\theta(n;x',n') - \theta(n;x',n')| = \left|\int_0^t P_s\mathcal{A}\theta(n;x',n')\right| \leq t \sup_{\mathbb{R}^d\times\bar{B}_\kappa} |\mathcal{A}\nabla_x\varphi| \|H^\varepsilon\|_{\mathrm{BM}(\mathbb{R}^d\times\bar{B}_\kappa)}.$$

This concludes the proof of the proposition. □

6.3 Perturbed test function method

Let $\varphi \in C_b^3(\mathbb{R}^d)$: φ is of class C^3 on $\mathbb{R}^d$ and φ with its derivatives up to order three are bounded on $\mathbb{R}^d$ (and φ is independent on n). By Theorem 5.8 and Section 6.2.2, we know that

$$\varphi(X_t^\varepsilon) - \varphi(x) - \int_0^t \mathcal{L}^\varepsilon\varphi(X_s^\varepsilon)ds \tag{6.48}$$

is a $(\mathcal{F}_t^\varepsilon)$-martingale. If (X_t^ε) converges (convergence on law is sufficient) to a certain Markov process (X_t) with generator $\mathcal{L}$ (we call $\mathcal{L}$ the *limit generator*), then, for φ possibly more regular, and in virtue of the Martingale characterization of Theorem 5.8

$$\varphi(X_t) - \varphi(x) - \int_0^t \mathcal{L}\varphi(X_s)ds \tag{6.49}$$

is a $(\mathcal{F}_t^X)$-martingale. We expect therefore the convergence of the set of equations $(0 \leq s \leq t)$

$$\mathbb{E}\left[\varphi(X_t^\varepsilon) - \varphi(X_s^\varepsilon) - \int_s^t \mathcal{L}^\varepsilon\varphi(X_\sigma^\varepsilon)d\sigma \middle| \mathcal{F}_t^\varepsilon\right] = 0, \tag{6.50}$$

to the set of equations

$$\mathbb{E}\left[\varphi(X_t) - \varphi(X_s) - \int_s^t \mathcal{L}\varphi(X_\sigma)d\sigma \middle| \mathcal{F}_t\right] = 0, \tag{6.51}$$

where $(\mathcal{F}_t)$ may be larger than $(\mathcal{F}_t^X)$. Although it gives, at least formally, the convergence (6.50)→(6.51), it is not reasonable to expect $\mathcal{L}^\varepsilon\varphi \to \mathcal{L}\varphi$. Indeed, generic test functions for $\mathcal{L}^\varepsilon$ depend on x and n. It is the approximation $\mathbb{E}\varphi(X_t^\varepsilon, m_t^\varepsilon) \simeq \mathbb{E}\bar{\varphi}(X_t)$, $\bar{\varphi}(x) := \langle\varphi(x,\cdot),\nu\rangle$ that induces a dependence solely on x at the limit. The idea of the perturbed test function method, devised (in our particular context) by Papanicolaou, Stroock and Varadhan, [PSV77], is to look for an expansion $\varphi^\varepsilon = \varphi + \varepsilon\varphi_1 + \varepsilon^2\varphi_2$ such that

$$\mathcal{L}^\varepsilon\varphi^\varepsilon = \mathcal{L}\varphi + o(1). \tag{6.52}$$

By identification of the powers of ε, (6.52) gives the following equations: at the order ε^{-2}, we have $\mathcal{A}\varphi = 0$, which is satisfied since φ is independent on n (actually this is an equivalence by Remark 6.1). At the order ε^{-1}, we obtain the equation

$$\mathcal{A}\varphi_1(n,x) + g(n,x)\cdot\nabla_x\varphi(x) = 0. \tag{6.53}$$

By (6.8) and (6.15), (6.53) has the solution

$$\varphi_1(\cdot, x) = R_0(g(\cdot, x)) \cdot \nabla_x \varphi(x), \tag{6.54}$$

which is an admissible test function. At the order ε^0, (6.52) gives the equation

$$\mathcal{A}\varphi_2(n, x) + g(n, x) \cdot \nabla_x \varphi_1(n, x) + f(x) \cdot \nabla_x \varphi(x) = \mathcal{L}\varphi(x). \tag{6.55}$$

Since $\langle \mathcal{A}\varphi_2(\cdot, x), \nu \rangle = 0$, a necessary condition to (6.55) is that

$$\mathcal{L}\varphi(x) = f(x) \cdot \nabla_x \varphi(x) + \langle g(\cdot, x) \cdot \nabla_x \varphi_1(\cdot, x), \nu \rangle. \tag{6.56}$$

The equation (6.56) gives the expression of the limit generator $\mathcal{L}$. We have

$$\nabla_x \varphi_1(x, \cdot) = \nabla_x R_0(g(\cdot, x)) \cdot \nabla_x \varphi(x)) + D_x^2 \varphi(x) \cdot R_0(g(\cdot, x)). \tag{6.57}$$

The limit generator is therefore

$$\mathcal{L}\varphi(x) = F(x) \cdot \nabla_x \varphi(x) + G(x){:}D^2 \varphi(x), \tag{6.58}$$

where

$$F(x) = f(x) + \langle g(\cdot, x) \cdot \nabla_x R_0(g(\cdot, x)), \nu \rangle, \quad G(x) = \langle g(\cdot, x) \otimes R_0(g(\cdot, x)), \nu \rangle. \tag{6.59}$$

In (6.58) and (6.59), we have used the following notations: $A{:}B$ is the canonical scalar product of two $d \times d$ matrices:

$$A{:}B = \sum_{i,j=1}^{d} a_{ij} b_{ij}. \tag{6.60}$$

If u, v are two vectors of $\mathbb{R}^d$, $u \otimes v$ is the rank-one $d \times d$ matrix defined by

$$(u \otimes v)_{ij} = u_i v_j. \tag{6.61}$$

Once $\mathcal{L}$ is defined by (6.56), we solve (6.55) by setting

$$\varphi_2 = R_0(\psi_2 - \langle \psi_2, \nu \rangle), \quad \psi_2(x, n) = g(n, x) \cdot \nabla_x \varphi_1(n, x) + f(x) \cdot \nabla_x \varphi(x). \tag{6.62}$$

Proposition 6.6 (Correctors). *Let $\varphi \in C_b^3(\mathbb{R}^d)$. Let φ_1 be defined by (6.54), let φ_2 be defined by (6.62). Then φ_1 and φ_2 are admissible test-functions in the sense of Definition 6.2 and the perturbed test function $\varphi^\varepsilon + \varphi + \varepsilon\varphi_1 + \varepsilon^2 \varphi_2$ satisfies*

$$|\mathcal{L}^\varepsilon \varphi^\varepsilon(x, n) - \mathcal{L}\varphi(x, n)\| \leq C(1 + |n|)\varepsilon, \quad \forall (x, n) \in \mathbb{R}^d \times \mathbb{R}^k, \tag{6.63}$$

where the constant C depends on f, g, φ, but not on ε, x, n.

Proof of Proposition 6.6. By Proposition 6.1 and Remark 6.2, φ_1 and φ_2 are admissible test-functions in the sense of Definition 6.2 and we have

$$\mathcal{L}^\varepsilon \varphi^\varepsilon - \mathcal{L}\varphi = \varepsilon(f \cdot \nabla_x \varphi_1 + g \cdot \nabla_x \varphi_2) + \varepsilon^2 f \cdot \nabla_x \varphi_2.$$

The bound (6.63) follows. $\square$

6.4 Tightness

In this section, we will show that (X_t^ε) is tight.

Proposition 6.7 (Tightness)**.** *Assume that m is almost surely continuous and satisfies (6.8), (6.9), (6.10). Assume that g satisfies*

$$R_0(g_i(\cdot,x)),\ |R_0(g_i(\cdot,x))|^2\in\mathcal{D}(\mathcal{A}),\quad A|R_0(g_i)|^2\in\mathrm{BM}(\bar{B}_\kappa\times\mathbb{R}^d),\tag{6.64}$$

for all $i\in\{1,\dots,d\}$, for all $x\in\mathbb{R}^d$. Let $T>0$. Then there exists $\alpha\in(0,1)$ and $C\geq 0$ independent on ε, such that, up to modification, the solution (X_t^ε) to (6.4)-(6.5) satisfies

$$X_t^\varepsilon=Y_t^\varepsilon+\zeta_t^\varepsilon,\quad \mathbb{E}\|Y^\varepsilon\|_{C^\alpha([0,T];\mathbb{R}^d)}\leq C,\quad \mathbb{E}\left[\sup_{t\in[0,T]}|\zeta_t^\varepsilon|\right]\leq C\varepsilon.\tag{6.65}$$

In particular, (X^ε) is tight in $C([0,T];\mathbb{R}^d)$.

Proof of Proposition 6.7. The last statement says, more exactly, that *the law of* (X^ε) is tight on $C([0,T];\mathbb{R}^d)$. This last assertion is a simple consequence of the bound (6.65). Indeed, (6.65) and the Markov inequality show that

$$\nu^\varepsilon(K_R^c)=\mathbb{P}(\|Y^\varepsilon\|_{C^\alpha([0,T];\mathbb{R}^d)}>R)\leq\frac{C}{R},$$

where $\nu^\varepsilon=\mathrm{Law}(X^\varepsilon)$, $K_R=\{Y\in C([0,T];\mathbb{R}^d);\|Y\|_{C^\alpha([0,T];\mathbb{R}^d)}\leq R\}$. By Ascoli's Theorem, K_R is compact. If $\eta>0$ is given, we have therefore $\nu^\varepsilon(K)\geq 1-\eta$ for all ε, where K is the compact $K_{C\eta^{-1}}$. This shows that (Y_t^ε) is tight. By Lemma 2.12 and Prohorov's Theorem (Theorem 2.13), (X_t^ε) is tight. Remember that

$$\|Y\|_{C^\alpha([0,T];\mathbb{R}^d)}=\sup_{t\in[0,T]}|Y(t)|+\sup_{s\neq t\in[0,T]}\frac{|Y(t)-Y(s)|}{|t-s|^\alpha}.$$

A consequence of (6.65), if $\varepsilon\leq 1$, is that $\mathbb{E}|X_t^\varepsilon|\leq 2C$ for all $t\in[0,T]$. Even this estimate is non-trivial. The right-hand side of (6.4) is singular, owing to the factor ε^{-1} and none of the classical techniques for ODEs, like the ones using Gronwall's lemma applied for example in the proof of Proposition 6.2, will give an estimate independent on ε. To obtain such an estimate, we will apply a perturbed test-function method at order 1. Let $\varphi\in C_b^2(\mathbb{R}^d)$ and let φ_1 be defined by (6.54): $\varphi_1=R_0(g)\cdot\nabla_x\varphi$. Set $\varphi^\varepsilon=\varphi+\varepsilon\varphi_1$. By Theorem 5.8 and Proposition 5.9,

$$M_t^\varepsilon:=\varphi^\varepsilon(X_t^\varepsilon,\bar{m}_t^\varepsilon)-\varphi(x,\bar{n})-\int_0^t\mathcal{L}^\varepsilon\varphi^\varepsilon(X_s^\varepsilon,\bar{m}_s^\varepsilon)ds\tag{6.66}$$

is a $(\mathcal{F}_t^\varepsilon)$-martingale with quadratic variation

$$\langle M^\varepsilon,M^\varepsilon\rangle_t=\int_0^t(\mathcal{L}^\varepsilon|\varphi^\varepsilon|^2-2\varphi^\varepsilon\mathcal{L}^\varepsilon\varphi^\varepsilon)(X_s^\varepsilon,\bar{m}_s^\varepsilon)ds.\tag{6.67}$$

Indeed, $|\varphi^\varepsilon|^2$ is in the domain of $\mathcal{L}^\varepsilon$. Thus is due to (6.64) and to the fact that the first-order terms (those in $\nabla_x\varphi$ in (6.43)) have no contribution in (6.67). It results that

$$|\langle M^\varepsilon, M^\varepsilon\rangle_t - \langle M^\varepsilon, M^\varepsilon\rangle_s| = \left|\int_s^t (\mathcal{A}|\varphi_1|^2 - 2\varphi_1\mathcal{A}\varphi_1)(X^\varepsilon_\sigma, \bar{m}^\varepsilon_\sigma)d\sigma\right| \leq C_1\|\nabla_x\varphi\|^2_{C_b(\mathbb{R}^d)}(t-s), \tag{6.68}$$

for $s \leq t$, where C_1 (and $C_2, \ldots$ in what follows) is a constant that may depend on x, g, f, γ_{mix} and T, but is independent on ε *and* φ (indeed, we will have to be careful to the dependence of our estimates upon φ, since at the end, we will take for φ the element of a sequence (φ_k) on $C^2_b(\mathbb{R}^d)$ converging to $x \mapsto x$). By the Burkholder-Davis-Gundy inequality, we deduce that

$$\mathbb{E}\left[|M^\varepsilon_t - M^\varepsilon_s|^4\right] \leq C_2\|\nabla_x\varphi\|^4_{C_b(\mathbb{R}^d)}(t-s)^2. \tag{6.69}$$

Admit for the moment that

$$|\mathcal{L}^\varepsilon\varphi^\varepsilon(y,n)| \leq C_3(\|\nabla_x\varphi\|_{C_b(\mathbb{R}^d)} + \|D^2\varphi\|_{C_b(\mathbb{R}^d)}), \tag{6.70}$$

for all $(y,n) \in \mathbb{R}^d \times \bar{B}_\kappa$. Then

$$\left|\int_s^t \mathcal{L}^\varepsilon\varphi^\varepsilon(X^\varepsilon_\sigma, \bar{m}^\varepsilon_\sigma)d\sigma\right|^4 \leq C_4\left(\|\nabla_x\varphi\|_{C_b(\mathbb{R}^d)} + \|D^2\varphi\|_{C_b(\mathbb{R}^d)}\right)^4(t-s)^4, \tag{6.71}$$

and, using the definition (6.66) and (6.69), we obtain

$$\mathbb{E}\left[|\varphi^\varepsilon(X^\varepsilon_t, \bar{m}^\varepsilon_t) - \varphi^\varepsilon(X^\varepsilon_s, \bar{m}^\varepsilon_s)|^4\right] \leq C_5(\|\nabla_x\varphi\|^4_{C_b(\mathbb{R}^d)} + \|D^2\varphi\|^4_{C_b(\mathbb{R}^d)})(t-s)^2. \tag{6.72}$$

The estimate (6.70) follows from the identities

$$\mathcal{L}^\varepsilon\varphi^\varepsilon = f\cdot\nabla\varphi + (g+\varepsilon f)\cdot\nabla\varphi_1, \quad \nabla\varphi_1 = R_0(\nabla_x g)\cdot\nabla_x\varphi + D^2\varphi\cdot R_0(g).$$

Let us now define the odd function φ_k by

$$\varphi_k(y) = \int_0^{y_i}\min(1, k^{-1}(z-2k)^-)dz, \quad y_i \geq 0. \tag{6.73}$$

The function φ_k is not C^2_b but $W^{2,\infty}$, which is enough for the validity of (6.72). We have $\varphi_k(y) \to y_i$ for all $y \in \mathbb{R}^d$ with $\|\nabla_x\varphi\|_{L^\infty(\mathbb{R}^d)} \leq 1$, $\|D^2\varphi\|_{L^\infty(\mathbb{R}^d)} \leq 1$. Therefore, we can take the limit in (6.72) applied to φ_k. We see that, if we set

$$\zeta^\varepsilon_t = \varepsilon R_0(g)(X^\varepsilon_t, \bar{m}^\varepsilon_t), \quad Y^\varepsilon_t = X^\varepsilon_t - \zeta^\varepsilon_t, \tag{6.74}$$

then we have

$$\mathbb{E}|Y^\varepsilon_t - Y^\varepsilon_s|^4 \leq C_6(t-s)^2. \tag{6.75}$$

At $t = 0$, we have $X^\varepsilon_0 = x$, $\zeta^\varepsilon_0 = \varepsilon R_0(g)(x,\bar{n})$, which is bounded. It follows that $\mathbb{E}|Y^\varepsilon_0| \leq C_7$. Using (3.26) and (6.75), we obtain (up to modification), the estimate $\mathbb{E}\|Y^\varepsilon\|_{C^\alpha([0,T];\mathbb{R}^d)} \leq C_8$, where $\alpha < \frac{1}{4}$. The estimate $\mathbb{E}\|\zeta^\varepsilon\|_{C([0,T];\mathbb{R}^d)} \leq C_9\varepsilon$ is clear since $R_0(g)$ is bounded. □

Proposition 6.8 (Tightness, càdlàg case). *Under the hypotheses of Proposition 6.7, save for the continuity of the process, we have the following result. The solution (X_t^ε) to (6.4)-(6.5) admits a decomposition $X_t^\varepsilon = Y_t^\varepsilon + \zeta_t^\varepsilon$ as the sum of two càdlàg processes (Y_t^ε) and (ζ_t^ε) where (Y_t^ε) is tight in the Skorohod space $D([0,T];\mathbb{R}^d)$ and (Y_t^ε) and (ζ_t^ε) satisfies the bound*

$$\mathbb{E}\left[\sup_{t\in[0,T]}|\zeta_t^\varepsilon|\right] \leq C\varepsilon, \quad \mathbb{E}\left[\sup_{t\in[0,T]}|Y_t^\varepsilon|\right] \leq C. \tag{6.76}$$

As a consequence, the family (X^ε) is tight in $C([0,T];\mathbb{R}^d)$.

Proof of Proposition 6.8. Note that considerations on the Skorohod topology on the space of càdlàg processes $D([0,T];\mathbb{R}^d)$ have not been introduced here before. We will give the necessary references from [Bil99] and [JS03]. The decomposition $X_t^\varepsilon = Y_t^\varepsilon + \zeta_t^\varepsilon$ is the same as in the proof of Proposition 6.7 (*cf.* (6.74)). We refer to this proof thus. First, the estimate (6.76) is straightforward. The estimate (6.71) on the increment on the integral term in (6.66) holds true, but the estimate on the martingale term (6.69) does not, since we cannot apply the Burkholder-Davis-Gundy inequality here. To prove that (Y_t^ε) is tight in $D([0,T];\mathbb{R}^d)$, we will apply the Aldous' criterion, [JS03, Theorem 4.5, p.356]. By Remark 5.7, we know that $|M_t^\varepsilon|^2 - A_t^\varepsilon$ is a martingale, where

$$A_t^\varepsilon = \int_0^t (\mathcal{L}^\varepsilon|\varphi^\varepsilon|^2 - 2\varphi^\varepsilon\mathcal{L}^\varepsilon\varphi^\varepsilon)(X_s^\varepsilon, \bar{m}_s^\varepsilon)ds.$$

Let $1 > \theta > 0$. Let τ_1, τ_2 be some $(\mathcal{F}_t^\varepsilon)$-stopping times such that

$$\tau_1 \leq \tau_2 \leq \tau_1 + \theta \text{ a.s.}, \quad \tau_2 \leq N, \tag{6.77}$$

for a given constant N. By the Doob optional sampling theorem, (5.9), we have

$$\mathbb{E}\left[|M_{\tau_2}^\varepsilon - M_{\tau_1}^\varepsilon|^2\right] = \mathbb{E}\left[|M_{\tau_2}^\varepsilon|^2 - |M_{\tau_1}^\varepsilon|^2\right].$$

Since $|M_t^\varepsilon|^2 - A_t^\varepsilon$ is a martingale, we deduce that

$$\mathbb{E}\left[|M_{\tau_2}^\varepsilon - M_{\tau_1}^\varepsilon|^2\right] = \mathbb{E}\left[A_{\tau_2}^\varepsilon - A_{\tau_1}^\varepsilon\right],$$

which gives $\mathbb{E}\left[|M_{\tau_2}^\varepsilon - M_{\tau_1}^\varepsilon|^2\right] \leq C\theta$. Similarly, (6.71) holds true when the terminal times are stopping times, hence

$$\mathbb{E}\left|\int_{\tau_1}^{\tau_2} \mathcal{L}^\varepsilon\varphi^\varepsilon(X_\sigma^\varepsilon, \bar{m}_\sigma^\varepsilon)d\sigma\right|^2 \leq C\left(\|\nabla_x\varphi\|_{C_b(\mathbb{R}^d)} + \|D^2\varphi\|_{C_b(\mathbb{R}^d)}\right)^2\theta^2. \tag{6.78}$$

We come back to the decomposition (6.66) to conclude that the increments of $\tilde{Y}_t^\varepsilon := \varphi(X_t^\varepsilon, \bar{m}_t^\varepsilon)$ satisfy the estimate $\mathbb{E}\left[|\tilde{Y}_{\tau_2}^\varepsilon - \tilde{Y}_{\tau_1}^\varepsilon|^2\right] \leq C\theta$. We then take a sequence (φ_k) as in (6.73) and let $k \to +\infty$ to obtain finally the bound $\mathbb{E}\left[|Y_{\tau_2}^\varepsilon - Y_{\tau_1}^\varepsilon|^2\right] \leq C\theta$ for possibly a different constant C. By the Markov inequality, this gives the property

$$\lim_{\theta\to 0}\limsup_{\varepsilon\in(0,1)}\sup_{\tau_1,\tau_2}\mathbb{P}(|Y_{\tau_2}^\varepsilon - Y_{\tau_1}^\varepsilon| > \eta) = 0$$

for all $\eta > 0$, $N > 0$, where, the sup on τ_1, τ_2 is, more precisely, the sup on stopping times satisfying (6.77). The Aldous' criterion (and the bound (6.76)) in being satisfied, (Y_t^ε) is tight in $D([0,T];\mathbb{R}^d)$. Using the estimate on the remainder ζ_t^ε in (6.76) and [JS03, Lemma 3.31 p.352], we deduce that (X_t^ε) is tight in $D([0,T];\mathbb{R}^d)$. Since (X_t^ε) is in $C([0,T];\mathbb{R}^d)$, it is actually tight in $C([0,T];\mathbb{R}^d)$. To establish this fact, one can use [Bil99, Theorem 13.2 p. 139] first, to deduce then, by [Bil99, (12.10) p. 123], that the condition [Bil99, (7.8) p. 82] on the standard modulus of continuity in $C([0,T];\mathbb{R}^d)$ is satisfied. The tightness of (X_t^ε) in $C([0,T];\mathbb{R}^d)$ then follows from [Bil99, Theorem 7.3 p. 82]. $\square$

6.5 The limit martingale problem

In this section, we will use the result of tightness established in Proposition 6.7 (or Proposition 6.8) to pass to the limit in the martingale characterization of (6.4)-(6.5). We refer to the discussion on the limit (6.50)→(6.51) at the beginning of Section 6.3. We consider (X_t^ε) for $\varepsilon \in \varepsilon_{\mathbb{N}}$, where $\varepsilon_{\mathbb{N}} = \{\varepsilon_n; n \in \mathbb{N}\}$ with $(\varepsilon_n) \downarrow 0$.

Proposition 6.9. *Assume that m satisfies (6.8), (6.9), (6.10). Assume that g satisfies (6.64). Then, up to subsequence, $(X^\varepsilon)_{\varepsilon\in\varepsilon_{\mathbb{N}}}$ is converging in law on $C([0,T];\mathbb{R}^d)$ to a process (X_t) satisfying the following martingale condition:*

$$\mathbb{E}\left[\varphi(X_t) - \varphi(X_s) - \int_s^t \mathcal{L}\varphi(X_\sigma)d\sigma \middle| \mathcal{F}_t^X\right] = 0, \tag{6.79}$$

for all $0 \leq s \leq t$ and for all $\varphi \in C_b^2(\mathbb{R}^d)$.

Proof of Proposition 6.9. Let $\varphi \in C_b^3(\mathbb{R}^d)$. Consider the modification $\varphi^\varepsilon = \varphi + \varepsilon\varphi_1 + \varepsilon^2\varphi_2$ to φ, with φ_1 and φ_2 defined by (6.54) and (6.62) respectively. We have

$$\mathbb{E}\left[\varphi^\varepsilon(X_t^\varepsilon, \bar{m}_t^\varepsilon) - \varphi^\varepsilon(X_s^\varepsilon, \bar{m}_s^\varepsilon) - \int_s^t \mathcal{L}^\varepsilon\varphi^\varepsilon(X_\sigma^\varepsilon, \bar{m}_\sigma^\varepsilon)d\sigma \middle| \mathcal{F}_t^\varepsilon\right] = 0, \tag{6.80}$$

for all $0 \leq s \leq t$. Let $m \in \mathbb{N}^*$, let $0 \leq t_1 < \cdots < t_m \leq t$ and let $\theta \in C_b(\mathbb{R}^m)$. Since (X_t^ε) is adapted, it follows from (6.80) that

$$\mathbb{E}\left[\left(\varphi^\varepsilon(X_t^\varepsilon, \bar{m}_t^\varepsilon) - \varphi^\varepsilon(X_s^\varepsilon, \bar{m}_s^\varepsilon) - \int_s^t \mathcal{L}^\varepsilon\varphi^\varepsilon(X_\sigma^\varepsilon, \bar{m}_\sigma^\varepsilon)d\sigma\right)\theta(X_{t_1}^\varepsilon, \ldots, X_{t_m}^\varepsilon)\right] = 0. \tag{6.81}$$

Using (6.63), we deduce from (6.81) that

$$\mathbb{E}\left[(\varphi(X_t^\varepsilon) - \varphi(X_s^\varepsilon))\theta(X_{t_1}^\varepsilon, \ldots, X_{t_m}^\varepsilon)\right] - \int_s^t \mathbb{E}\left[\mathcal{L}\varphi(X_\sigma^\varepsilon)\theta(X_{t_1}^\varepsilon, \ldots, X_{t_m}^\varepsilon)\right] d\sigma = \mathcal{O}(\varepsilon). \tag{6.82}$$

Up to a subsequence, $(X^\varepsilon)_{\varepsilon\in\varepsilon_\mathbb{N}}$ converges in law on $C([0,T];\mathbb{R}^d)$ to a stochastic process X on $C([0,T];\mathbb{R}^d)$. By taking the limit in (6.82) along this subsequence, we have

$$\mathbb{E}\left[\left(\varphi(X_t) - \varphi(X_s) - \int_s^t \mathcal{L}\varphi(X_\sigma)d\sigma\right)\theta(X_{t_1},\dots,X_{t_m})\right] = 0. \tag{6.83}$$

Since $C_b^3(\mathbb{R}^d)$ is π-dense in $C_b^2(\mathbb{R}^d)$, (6.83) holds true when $\varphi \in C_b^2(\mathbb{R}^d)$. Define then the following finite signed measure Q on Ω

$$Q(B) = \mathbb{E}\left[\left(\varphi(X_t) - \varphi(X_s) - \int_s^t \mathcal{L}\varphi(X_\sigma)d\sigma\right)\mathbf{1}_B\right].$$

We want to show that $Q(B) = 0$ for all $B \in \mathcal{F}_s^X$. Since $\mathcal{F}_s^X$ is generated by cylindrical sets, it is sufficient to show that $\mu(A) = 0$ for all $A \in \mathcal{B}(\mathbb{R}^m)$, where

$$\mu(A) = \mathbb{E}\left[\left(\varphi(X_t) - \varphi(X_s) - \int_s^t \mathcal{L}\varphi(X_\sigma)d\sigma\right)\mathbf{1}_A(X_{t_1},\dots,X_{t_m})\right].$$

We have $\langle\theta,\mu\rangle = 0$ for all $\theta \in C_b(\mathbb{R}^m)$, therefore $\mu(A) = 0$ since $C_b(\mathbb{R}^m)$ is a separating class (Proposition 2.10). This gives the desired result. □

6.6 Identification of the limit and conclusion

6.6.1 Auto-correlation function of a stationary stochastic process

The definition of stationary process has already been given (Definition 3.10). Here we focus on the definition and properties of the auto-correlation function of stationary processes. If H is a Hilbert space and $u, v \in H$, then $u \otimes v$ is the operator defined by

$$\langle u \otimes v \cdot x, y\rangle_H = \langle u, x\rangle_H \langle v, y\rangle_H$$

for all $x, y \in H$.

Definition 6.3 (Auto-correlation)**.** Let H be a separable Hilbert space. Let $(X_t)_{t\geq 0}$ be an H-valued process such that $\mathbb{E}\|X_t\|_H^2 < +\infty$ for all t. We assume $\mathbb{E}[X_t] = 0$ for all t (X is centred). The auto-correlation function of $(X_t)_{t\geq 0}$ is the operator $\Gamma(t,s)\colon H \to H$ defined by $\Gamma(t,s) = \mathbb{E}\left[X(s) \otimes X(t)\right]$.

Note that $\|\Gamma(t,s)\|_{H\to H} \leq \mathbb{E}\|X(s)\|_H\|X(t)\|_H$. If X is not centred, the definition should be modified into

$$\Gamma(t,s) = \mathbb{E}\left[(X(s) - \mathbb{E}[X(s)]) \otimes (X(t) - \mathbb{E}[X(t)])\right].$$

For a stationary process, the auto-correlation function depends on $t - s$ only, and we set

$$\Gamma(t) = \mathbb{E}\left[X(s) \otimes X(s+t)\right]. \tag{6.84}$$

Proposition 6.10. *Let H be a separable Hilbert space. Let $(X_t)_{t\geq 0}$ be an H-valued stationary process such that $\mathbb{E}\|X_t\|_H^2 < +\infty$ for all t and such that, for all $x \in H$, the map $t \mapsto \langle \Gamma(t)x, x\rangle$ is integrable on $\mathbb{R}_+$. Then the following integral is non-negative:*

$$\int_0^\infty \langle \Gamma(t)x, x\rangle dt \geq 0. \tag{6.85}$$

If the integral over $\mathbb{R}_+$ of $\Gamma(t)$ is convergent in the space of linear bounded operators on H, (6.85) asserts that $\int_0^\infty \Gamma(t)dt$ is a positive operator on H.

Proof of Proposition 6.10. We will use the following result.

Lemma 6.11. *Let E be a separable Banach space. Let $(X_t)_{t\geq 0}$ be an E-valued stationary process. Let $\theta \in L^1(\mathbb{R}_+)$ be the density of a probability measure and let $\varphi \in \mathrm{BM}(E)$. Then*

$$\mathbb{E}\left|\int_0^{+\infty} \theta(t)\varphi(X_t)dt\right|^2 = 2\int_0^{+\infty} \theta * \check{\theta}(t)\mathbb{E}\left[\varphi(X_0)\varphi(X_t)\right]dt, \tag{6.86}$$

where $\theta * \check{\theta}$ is defined by

$$\theta * \check{\theta}(t) := \int_0^{+\infty} \theta(t+s)\theta(s)ds, \tag{6.87}$$

for a.e. $t \geq 0$.

We admit Lemma 6.11 for the moment. Note that each term in (6.86) makes sense: indeed, the random variable

$$\int_0^{+\infty} \theta(t)\varphi(X_t)dt$$

is an average of $\varphi(X_t)$. The left-hand side of (6.86) is finite as soon as $\mathbb{E}|\varphi(X_t)|^2$ (which is independent on t) is finite. If we extend θ by 0 on $\mathbb{R}_-$ and define $\check{\theta}(s) = \theta(-s)$, then $\theta * \check{\theta}$ as defined in (6.87) is really a convolution product. In particular, $\theta * \check{\theta}$ is also a density probability and is well defined a.e. Consequently, we reach the same conclusion for the right-hand side of (6.86). It is well defined if $\mathbb{E}|\varphi(X_t)|^2$ is finite. These considerations prove that (6.86) can be extended, thanks to an argument of approximation, to $\varphi(X_t) = \langle X_t, x\rangle_H$. It gives us

$$\int_0^{+\infty} \theta * \check{\theta}(t)\langle \Gamma(t)x, x\rangle_H dt = \mathbb{E}\left|\int_0^{+\infty} \theta(t)\langle X_t, x\rangle_H dt\right|^2 \geq 0 \tag{6.88}$$

We apply (6.88) with $\theta(t) = \theta_\lambda(t) = \lambda e^{-\lambda t}\mathbf{1}_{t>0}$ for $\lambda > 0$ (this is the probability density of the exponential distribution with parameter λ). Computing first (for $t \geq 0$)

$$\theta_\lambda * \check{\theta}_\lambda(t) := \lambda^2 e^{-\lambda t}\int_0^{+\infty} e^{-2\lambda s}ds = \frac{1}{2}\theta_\lambda(t),$$

we obtain, after division by λ,

$$\int_0^{+\infty} e^{-\lambda t}\langle \Gamma(t)x, x\rangle_H dt = \lambda\mathbb{E}\left|\int_0^t e^{-\lambda t}\langle X_t, x\rangle_H dt\right|^2 \geq 0. \tag{6.89}$$

We let $\lambda \to 0$ to conclude. $\square$

Proof of Lemma 6.11. This is a simple computation. By the Fubini theorem, we have

$$\mathbb{E}\left|\int_0^{+\infty}\theta(t)\varphi(X_t)dt\right|^2=\int_0^{+\infty}\int_0^{+\infty}\theta(t)\theta(s)\mathbb{E}[\varphi(X_t)\varphi(X_s)]dsdt.$$

We make the distinction between the domains of integration $\{s\leq t\}$, $\{s\geq t\}$. By symmetry of the argument, we obtain

$$2\int_{t=0}^{+\infty}\int_{s=0}^{t}\theta(t)\theta(s)\mathbb{E}[\varphi(X_t)\varphi(X_s)]dsdt=2\int_{t=0}^{+\infty}\int_{s=0}^{t}\theta(t)\theta(s)\mathbb{E}[\varphi(X_{t-s})\varphi(X_0)]dsdt,$$

since $(X_t)_{t\geq 0}$ is stationary. We use the change of variable $s'=t-s$ and Fubini's theorem to get

$$2\int_{t=0}^{+\infty}\int_{s=0}^{t}\theta(t)\theta(t-s)\mathbb{E}[\varphi(X_s)\varphi(X_0)]dsdt=2\int_{s=0}^{+\infty}\int_{t=s}^{+\infty}\theta(t)\theta(s)\mathbb{E}[\varphi(X_s)\varphi(X_0)]dsdt,$$

which yields (6.86). □

6.6.2 Diffusion operator

Recall the formula (6.58)-(6.59) for the limit generator $\mathcal{L}$. By Proposition 6.10 applied with $H=\mathbb{R}^d$, the matrix $G(x)$ is symmetric and non-negative. Indeed, for $\xi\in\mathbb{R}^d$, and by the Fubini Theorem, we have

$$\langle G(x)\xi,\xi\rangle_{\mathbb{R}^d}=\mathbb{E}[\langle g(\bar{n},x),\xi\rangle_{\mathbb{R}^d}\langle R_0(g)(\bar{n},x),\xi\rangle_{\mathbb{R}^d}]=\int_0^\infty\mathbb{E}[P_t\varphi(\bar{n})\varphi(\bar{n})]dt,$$

where $\varphi(n):=\langle g(n,x),\xi\rangle_{\mathbb{R}^d}$ (x being fixed here). By the Markov property, we have, since $\bar{m}_t=\bar{n}$ at $t=0$,

$$\mathbb{E}[P_t\varphi(\bar{n})\varphi(\bar{n})]=\mathbb{E}\left[\mathbb{E}[\varphi(\bar{m}_t)|\mathcal{F}_0]\varphi(\bar{m}_0)\right]=\mathbb{E}\left[\mathbb{E}[\varphi(\bar{m}_t)\varphi(\bar{m}_0)|\mathcal{F}_0]\right]=\langle\Gamma_{g(x)}(t)\xi,\xi\rangle_{\mathbb{R}^d},$$

where $\Gamma_{g(x)}(t)$ is the autocorrelation function of $g(\bar{m}_t,x)$. Eventually, we obtain

$$\langle G(x)\xi,\xi\rangle_{\mathbb{R}^d}=\int_0^\infty\langle\Gamma_{g(x)}(t)\xi,\xi\rangle_{\mathbb{R}^d}dt\geq 0.$$

Set $X_t=g(\bar{m}_t,x)$. The proof of Proposition 6.10 also shows that $\langle G(x)\xi,\xi\rangle_{\mathbb{R}^d}$ is the limit when $\lambda\to 0$ of

$$\lambda\mathbb{E}\left|\int_0^\infty e^{-\lambda t}\langle X_t,\xi\rangle_{\mathbb{R}^d}dt\right|^2=\lambda\mathbb{E}\int_0^\infty\int_0^\infty e^{-\lambda(t+s)}\langle X_t,\xi\rangle_{\mathbb{R}^d}\langle X_s,\xi\rangle_{\mathbb{R}^d}dsdt=\langle G_\lambda(x)\xi,\xi\rangle,$$

where

$$G_\lambda(x):=\lambda\int_0^\infty\int_0^\infty e^{-\lambda(t+s)}\mathbb{E}[X_t\otimes X_s]dsdt.$$

Since $G_\lambda(x)$ is symmetric, $G(x)$ is symmetric. We will make the following hypothesis: we assume that there exists

$$\sigma \in \mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^{d\times d}), \quad G(x) = \sigma^*(x)\sigma(x), \quad \forall x \in \mathbb{R}^d. \tag{6.90}$$

The existence of such a regular square root of $G(x)$ is false in general (consider the case $G(x) = |x|\mathrm{Id}$). It is true if $G(x) \geq \alpha\mathrm{Id}$ for $\alpha > 0$, we can then take

$$\sigma(x) = \frac{1}{2\pi i}\int_C (z - G(x))^{-1}\sqrt{z}dz, \quad \sqrt{z} = \exp\left(\frac{1}{2}\log(z)\right),$$

where $\log(z)$ is the determination of the logarithm on $\mathbb{C}\backslash\mathbb{R}_-$ and C a circle with diameter $[z_1, z_2]$, where $z_1 = \frac{\alpha}{2}$, and $z_2 = 2M$ with M large enough to ensure $G(x) \leq M\mathrm{Id}$ for all x. Under (6.90), the generator $\mathcal{L}$ is the generator associated to the process Y_t solution the stochastic differential equation (SDE)

$$dY_t = F(Y_t)dt + \sigma(Y_t)dB_t, \tag{6.91}$$

with initial datum

$$Y_0 = y \in \mathbb{R}^d, \tag{6.92}$$

where (B_t) is a d-dimensional Wiener process. In the next sections, where we will study the stochastic integral and stochastic differential equations, we will show the following result.

Theorem 6.12. *Let $f\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be Lipschitz continuous functions. Then (6.91)-(6.92) has a unique solution $Y_t(y)$ in the space of adapted processes in $C([0,T];L^2(\Omega;\mathbb{R}^d))$. The process $(Y_t(y))$ is a Markov process on $\mathbb{R}^d$. Its generator generator $\mathcal{A}$ contains the unbounded generator $\mathcal{A}_0$, defined by:*

$$\mathcal{D}(\mathcal{A}_0) = \left\{\varphi \in C^2_b(\mathbb{R}^d);\, \sup_{x\in\mathbb{R}^d}\left[|x|\|D\varphi(x)\|_{\mathcal{L}(\mathbb{R}^d;\mathbb{R}^d)} + |x|^2_K\|D^2\varphi(x)\|_{\mathcal{L}(\mathbb{R}^d\times\mathbb{R}^d;\mathbb{R}^d)}\right] < +\infty\right\},$$

$$\mathcal{A}_0\varphi = \mathcal{L}\varphi,$$

where $\mathcal{L}$ is defined in (6.56).

Using Theorem 6.12, we obtain the following result.

Theorem 6.13 (Diffusion-approximation in finite dimension)**.** *Assume that m satisfies (6.8), (6.9), (6.10). Assume that g satisfies (6.64). Then $(X^\varepsilon)_{\varepsilon>0}$, the solution to (6.4)-(6.5), is converging in law on $C([0,T];\mathbb{R}^d)$ to the solution $(Y_t(x))$ to (6.91) with initial condition $Y_0(x) = x$.*

Remark 6.3 (Terminology)*.* The solution to a SDE like (6.91) is called a *diffusion*. Theorem 6.13 states that (X^ε_t), the solution to (6.4)-(6.5), can be approached (in law) by the diffusion $Y_t(x)$. This is a result of *diffusion-approximation*.

Proof of Theorem 6.13. Let $(\Pi^\sharp_t)_{t\geq 0}$ denote the semi-group of transition operators associated to $(Y_t(y))$. By[8] Theorem 5.10, we have

$$\mathbb{E}\Big[\varphi(X_{s+t})\Big|\mathcal{F}^X_t\Big]=\Pi^\sharp_s\varphi(X_t), \tag{6.93}$$

for all $s,t\geq 0$, for all $\varphi\in C_b(\mathbb{R}^d)$. Taking $t=0$ in (6.93), we see that $\Pi_s\varphi(x)=\Pi^\sharp_s\varphi(x)$ for all $\varphi\in C_b(\mathbb{R}^d)$, where $\Pi_s\varphi(x):=\mathbb{E}\varphi(X_s)$. This identity shows that the law of $X_s(x)$ and the law of $Y_s(x)$ coincide when tested against functions in $C_b(\mathbb{R}^d)$. Since $C_b(\mathbb{R}^d)$ is a separating class, they are identical. Consequently $\Pi_s=\Pi^\sharp_s$. As in the end of the proof of Proposition 6.3, we can show that (6.93) holds true when $\varphi\in\mathrm{BM}(\mathbb{R}^d)$. This shows that (X_t) is Markov. Since $(X_t(x))$ and $(Y_t(x))$ have same law at time $t=0$, it follows from Proposition 4.1 that the processes $(X_t(x))$ and $(Y_t(x))$ have same law. We have shown that $(X^\varepsilon_t)_{\varepsilon\in\varepsilon_\mathbb{N}}$ has a subsequence converging in law, and that the limit is uniquely determined: it is $Y_t(x)$. By uniqueness of the limit, the whole sequence is converging: $(X^\varepsilon_t)\to(Y_t(x))$ in law on $C([0,T];\mathbb{R}^d)$. □

7 Stochastic integration

Let $(\beta(t))$ be a one dimensional Wiener process over $(\Omega,\mathcal{F},\mathbb{P})$. Let K be a separable Hilbert space and let $(g(t))$ be a K-valued stochastic process. The first obstacle to the definition of the stochastic integral

$$I(g)=\int_0^T g(t)d\beta(t) \tag{7.1}$$

is the lack of regularity of $t\mapsto\beta(t)$, which has almost-surely a regularity $1/2-$: for all $\alpha\in[0,1/2)$, almost-surely, β is in $C^\alpha([0,T])$ and not in $C^{1/2}([0,T])$. Young's integration theory can be used to give a meaning to (7.1) for integrands $g\in C^\gamma([0,T])$ when $\gamma>1/2$, but this not applicable here, since the resolution of stochastic differential equation requires a definition of $I(\beta)$. In that context, one has to expand the theory of Young's or Riemann – Stieltjes' Integral, this is one of the purpose of rough paths' theory, *cf.* [FH14]. Below, it is the martingale properties of the Wiener process which are used to define the stochastic integral (7.1).

7.1 Stochastic integration of elementary processes

Let $(\mathcal{F}_t)_{t\geq 0}$ be a given filtration, such that $(\beta(t))$ is $(\mathcal{F}_t)$-adapted, and the increment $\beta(t)-\beta(s)$ is independent on $\mathcal{F}_s$ for all $0\leq s\leq t$. Let $(g(t))_{t\in[0,T]}$ be an K-valued stochastic process which is adapted, simple and L^2, in the sense that

$$g(\omega,t)=g_{-1}(\omega)\mathbf{1}_{\{0\}}(t)+\sum_{i=0}^{n-1}g_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t), \tag{7.2}$$

[8]since $\mathcal{A}$ may not be $\mathcal{L}$ (we do not want to investigate the domain of $\mathcal{L}$), we have to adapt slightly the proof of Theorem 5.10: we consider first test-functions $\varphi\in\mathcal{D}(\mathcal{A}_0)$, and obtain (6.93), then we use the fact that $\mathcal{D}(\mathcal{A}_0)$ is π-dense in $C_b(\mathbb{R}^d)$ to get the general result

where $0 \leq t_0 \leq \cdots \leq t_n \leq T$, g_{-1} is $\mathcal{F}_0$-measurable, each g_i, $i \in \{0, \dots, n-1\}$ is $\mathcal{F}_{t_i}$-measurable and in $L^2(\Omega; K)$. For such an integrand g, we define $I(g)$ as the following Riemann sum

$$I(g) = \sum_{i=0}^{n-1} (\beta(t_{i+1}) - \beta(t_i)) g_i. \tag{7.3}$$

Remark 7.1. Let λ denote the Lebesgue measure on $[0, T]$. For g as in (7.2), we have

$$g(\omega, t) = \sum_{i=0}^{n-1} g_i(\omega) \mathbf{1}_{(t_i, t_{i+1}]}(t),$$

for $\mathbb{P} \times \lambda$-almost all $(\omega, t) \in \Omega \times [0, T]$ since the singleton $\{0\}$ has λ-measure 0. We include the term $g_{-1}(\omega)\mathbf{1}_{\{0\}}(t)$ in (7.2) to be consistent with the definition of the predictable σ-algebra in the next section 7.2. Consistency here is in the sense that the predictable σ-algebra $\mathcal{P}_T$ as defined in Section 7.2 is precisely the σ-algebra generated by the elementary processes.
Note that g as in (7.2) belongs to $L^2(\Omega \times [0, T], \mathbb{P} \times \lambda)$ and that

$$\int_0^T \mathbb{E} \|g(t)\|_K^2 dt = \sum_{i=0}^{n-1} (t_{i+1} - t_i) \mathbb{E} \left[\|g_i\|_K^2 \right]. \tag{7.4}$$

In (7.3), g_i and the increment $\beta(t_{i+1}) - \beta(t_i)$ are independent. Using this fact, we can prove the following proposition.

Proposition 7.1 (Itō's isometry). *We have $I(g) \in L^2(\Omega; K)$ and*

$$\mathbb{E}[I(g)] = 0, \quad \mathbb{E}\left[\|I(g)\|_K^2\right] = \int_0^T \mathbb{E}\|g(t)\|_K^2 dt. \tag{7.5}$$

Proof of Proposition 7.1. We develop the square of the norm of $I(g)$:

$$\begin{aligned} \|I(g)\|_K^2 = \sum_{i=0}^{n-1} &|\beta(t_{i+1}) - \beta(t_i)|^2 \|g_i\|_K^2 \\ &+ 2 \sum_{0 \leq i < j \leq n-1} (\beta(t_{i+1}) - \beta(t_i))(\beta(t_{j+1}) - \beta(t_j)) \langle g_i, g_j \rangle_K. \end{aligned} \tag{7.6}$$

By independence, the expectancy of the second term (cross-products) in (7.6) vanishes, while the expectancy of the first term gives

$$\sum_{i=0}^{n-1} (t_{i+1} - t_i) \mathbb{E}\left[\|g_i\|_E^2\right] = \int_0^T \mathbb{E}\|g(t)\|_E^2 dt$$

since $\mathbb{E}\left[|\beta(t_{i+1}) - \beta(t_i)|^2\right] = (t_{i+1} - t_i)$. This shows that $I(g) \in L^2(\Omega; K)$ and the second equality in (7.5). The first equality follows from the identity

$$\mathbb{E}[(\beta(t_{i+1}) - \beta(t_i)) g_i] = \mathbb{E}[(\beta(t_{i+1}) - \beta(t_i))] \, \mathbb{E}[g_i] = 0,$$

for all $i \in \{0, \dots, n-1\}$. $\square$

7.2 Extension

Let $\mathcal{E}_T$ denote the set of L^2-elementary predictable functions in the form (7.2). This is a subset of $L^2(\Omega \times [0,T]; K)$ (the measure on $\Omega \times [0,T]$ being the product measure $\mathbb{P} \times \lambda$). The second identity in (7.5) shows that

$$I \colon \mathcal{E}_T \subset L^2(\Omega \times [0,T]; K) \to L^2(\Omega; K) \tag{7.7}$$

is a linear isometry. The stochastic integral $I(g)$ is the extension of this isometry to the closure $\overline{\mathcal{E}_T}$ of $\mathcal{E}_T$ in $L^2(\Omega \times [0,T]; K)$. It is clear that (7.5) (Itō's isometry) is preserved in this extension operation. To understand what is $I(g)$ exactly, we have to identify the closure $\overline{\mathcal{E}_T}$, or, at least certain sub-classes of $\overline{\mathcal{E}_T}$. For this purpose, we introduce $\mathcal{P}_T$, the predictable sub-σ-algebra of $\mathcal{F} \times \mathcal{B}([0,T])$ generated by the sets $F_0 \times \{0\}$, $F_s \times (s,t]$, where F_0 is $\mathcal{F}_0$-measurable, $0 \leq s < t \leq T$ and F_s is $\mathcal{F}_s$-measurable. We have denoted by $\mathcal{B}([0,T])$ the Borel σ-algebra on $[0,T]$. It is clear that each element in $\mathcal{E}_T$ is $\mathcal{P}_T$ measurable. We will admit without proof the following propositions (Proposition 7.2 and Proposition 7.3).

Proposition 7.2. *Assume that the filtration $(\mathcal{F}_t)$ is complete and continuous from the right. Then the σ-algebra generated on $\Omega \times [0,T]$ by adapted left-continuous (respectively, adapted continuous processes) coincides with the predictable σ-algebra $\mathcal{P}_T$.*

Proof of Proposition 7.2. Exercise, or see [RY99, Proposition 5.1, p. 171]. □

A $\mathcal{P}_T$-measurable process is called a predictable process. Denote by $\mathcal{P}_T^*$ the completion of $\mathcal{P}_T$. By Proposition 7.2, any adapted a.s. left-continuous or continuous process is $\mathcal{P}_T^*$-measurable.

Proposition 7.3. *Assume that the filtration $(\mathcal{F}_t)$ is complete and continuous from the right. Define*

1. *the optional σ-algebra to be the σ-algebra $\mathcal{O}$ generated by adapted càdlàg processes,*
2. *the progressive σ-algebra to be the σ-algebra* Prog *generated by the progressively measurable processes (Definition 4.14).*

Then we have the inclusion

$$\mathcal{P}_T \subset \mathcal{O} \subset \mathrm{Prog} \subset \mathcal{P}_T^*, \tag{7.8}$$

and the identity

$$\overline{\mathcal{E}_T} = L^2(\Omega \times [0,T], \mathcal{P}_T^*; K). \tag{7.9}$$

Proof of Proposition 7.3. See [CW90, Lemma 2.4] and [CW90, Chapter 3]. □

In what follows we will always assume that the filtration $(\mathcal{F}_t)$ is complete and continuous from the right.

Note that a function is in $L^2(\Omega \times [0,T], \mathcal{P}_T^*; K)$ if it is equal $\mathbb{P} \times \lambda$-a.e. to a function of $L^2(\Omega \times [0,T]; K)$ which is $\mathcal{P}_T$-measurable.

A consequence of Proposition 7.2 and Proposition 7.3 is that we can define the stochastic integral $I(g)$ of processes $(g(t))$ which are either adapted and left-continuous or continuous or càdlàg or progressively measurable. We will use the notation $\int_0^T g(t)d\beta(t)$ for $I(g)$.

Exercise 7.1. Show that (in the case $K = \mathbb{R}$)

1. if $(g(t))$ is an adapted process such $g \in C([0,T]; L^2(\Omega))$, then

$$\int_0^T g(t)d\beta(t) = \lim_{|\sigma|\to 0} \sum_{i=0}^{n-1} g(t_i)(\beta(t_{i+1}) - \beta(t_i)), \tag{7.10}$$

   where $\sigma = \{0 = t_0 \leq \cdots \leq t_n = T\}$ and $\sigma = \sup_{0\leq i<n}(t_{i+1} - t_i)$.

2. Show that the result (7.10) holds true if $(g(t))$ is a continuous adapted process such that $\sup_{t\in[0,T]} \mathbb{E}|g(t)|^q$ is finite for a $q > 2$.

3. If $g \in L^2(0,T)$ is deterministic, then $\int_0^T g(t)d\beta(t)$ is a gaussian random variable $\mathcal{N}(0,\sigma^2)$ of variance

$$\sigma^2 = \int_0^t |g(t)|^2 dt.$$

The solution to Exercise 7.1 is here.

7.3 Continuity and martingale property

Lemma 7.4 (Conditional Itô's isometry)**.** *Assume that the filtration $(\mathcal{F}_t)$ is complete and continuous from the right. Then the identity*

$$\mathbb{E}\left[\left|\int_s^t g(\sigma)d\beta(\sigma)\right|^2 \mid \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t |g(\sigma)|^2 \, d\sigma \mid \mathcal{F}_s\right], \tag{7.11}$$

is satisfied for all $g \in L^2(\Omega \times [0,T], \mathcal{P}_T^*; \mathbb{R})$ and all $0 \leq s \leq t \leq T$.

The proof of (7.11) is a variant of the proof of the Itô's isometry. We will use Lemma 7.4 to compute the quadratic variation of the stochastic integral.

Proposition 7.5. *Assume that the filtration $(\mathcal{F}_t)$ is complete and continuous from the right. Let $g \in L^2(\Omega \times [0,T], \mathcal{P}_T^*; \mathbb{R})$. Then the stochastic integral*

$$M(g)_t = \int_0^t g(s)d\beta(s) \tag{7.12}$$

is a continuous $(\mathcal{F}_t)$-martingale with quadratic variation

$$\langle M(g), M(g)\rangle_t = \int_0^t |g(s)|^2 ds. \tag{7.13}$$

Proof of Proposition 7.5. Let (g_n) be a sequence of elementary predictable functions that converges to g in $L^2(\Omega \times [0,T])$. By the Itô isometry, we have

$$\mathbb{E}|M(g_n)_t - M(g)_t|^2 = \int_0^t |g_n(s) - g(s)|^2 ds \to 0, \tag{7.14}$$

for every $t \in [0,T]$. Before we begin the study of $(M(g)_t)$, let us remark that we have the consistency relation

$$M(g)_t = \int_0^T \mathbf{1}_{[0,t]} g(s) ds. \tag{7.15}$$

Indeed, $(\mathbf{1}_{[0,t]} g_n)$ is a sequence of elementary predictable functions that converges to $\mathbf{1}_{[0,t]} g$ in $L^2(\Omega \times [0,T])$ and if

$$g = \sum_{i=0}^{n-1} g_i \mathbf{1}_{(t_i, t_{i+1}]}$$

is simple, then

$$\mathbf{1}_{[0,t]} g = \sum_{i=0}^{n-1} g_i \mathbf{1}_{(t\wedge t_i, t\wedge t_{i+1}]},$$

hence

$$M(\mathbf{1}_{[0,t]} g)_T = \sum_{i=0}^{n-1} g_i(\beta(t\wedge t_{i+1}) - \beta(t \wedge t_i)) = M(g)_t, \tag{7.16}$$

since $(t \wedge t_i)_{0,n}$ is a subdivision of $[0,t]$ according to which $g|_{[0,t]}$ has a decomposition of elementary predictable function. If g is elementary, (7.16) shows that $(M(g)_t)_{t\in[0,T]}$ is a continuous $(\mathcal{F}_t)$-martingale. By the Doob inequality (5.13) and the Itô isometry, we have the bound

$$\begin{aligned}\mathbb{E} \sup_{t\in[0,T]} |M(g_n)_t - M(g_p)_t|^2 &= \mathbb{E} \sup_{t\in[0,T]} |M(g_n - g_p)_t|^2 \\ &\leq 4\mathbb{E}|M(g_n - g_p)_T|^2 = 4\mathbb{E}\int_0^T |g_n - g_p|^2 dt, \end{aligned} \tag{7.17}$$

for $n, p \geq 0$. The estimate (7.17) shows that $(M(g_n)_t)$ satisfies a Cauchy condition in the complete space $F = L^2(\Omega; C([0,T];\mathbb{R}))$. Consequently, the sequence $(M(g_n)_t)$ is convergent in this space, and since convergence in F implies the simple convergence (7.14), the limit is $(M(g)_t)$. This shows that $(M(g))_t$ is a continuous martingale. To compute the quadratic variation of $(M(g))_t$, we use Proposition 5.6: let $\sigma=(t_i)_{0,n}$ be a subdivision of $[0,T]$. By (7.11) we have

$$\bar{V}_\sigma^{(2)}(t):=\sum_{i=0}^{n-1}\mathbb{E}\left[|M(g)_{t\wedge t_{i+1}}-M(g)_{t\wedge t_i}|^2|\mathcal{F}_{t_i}\right]=\sum_{i=0}^{n-1}\mathbb{E}\left[\int_{t\wedge t_i}^{t\wedge t_{i+1}}|g(s)|^2ds|\mathcal{F}_{t_i}\right]. \quad (7.18)$$

From (7.18), we deduce that, at fixed time t, $g\mapsto\langle M(g),M(g)\rangle_t$ is continuous $L^2(\Omega)\to L^1(\Omega)$. Indeed, if $\bar{V}_\sigma^{(2)}(t)$ and $\bar{W}_\sigma^{(2)}(t)$ are the discrete quadratic variation associated to the integrands g and h respectively, then

$$\begin{aligned}\mathbb{E}\left|\bar{V}_\sigma^{(2)}(t)-\bar{W}_\sigma^{(2)}(t)\right| &\leq\sum_{i=0}^{n-1}\int_{t\wedge t_i}^{t\wedge t_{i+1}}\mathbb{E}||g(s)|^2-|h(s)|^2|ds=\int_0^t\mathbb{E}||g(s)|^2-|h(s)|^2|ds\\ &\leq\left[\int_0^t\mathbb{E}|g(s)-h(s)|^2ds\right]^{1/2}\left[\int_0^t\mathbb{E}|g(s)+h(s)|^2ds\right]^{1/2}.\end{aligned}$$

Consequently, we may assume without loss of generality that g is bounded. Then we can prove that

$$(7.18)=\sum_{i=0}^{n-1}\int_{t\wedge t_i}^{t\wedge t_{i+1}}|g(s)|^2ds+o(1)=\int_0^t|g(s)|^2ds+o(1),$$

when $[|\sigma|\to 0]$, where the $o(1)$ is in $L^2(\Omega)$ (same proof as Step 1. of the proof of Proposition 5.9, *cf.* (5.51)). This gives the result. □

7.4 Itô's Formula

Proposition 7.6 (Itô's Formula)**.** *Assume that the filtration $(\mathcal{F}_t)$ is complete and continuous from the right. Let $g\in L^2(\Omega\times[0,T],\mathcal{P}_T^*;\mathbb{R})$, $f\in L^1(\Omega\times[0,T],\mathcal{P}_T^*;\mathbb{R})$, let $x\in\mathbb{R}$ and let*

$$X_t=x+\int_0^t f(s)ds+\int_0^t g(s)d\beta(s).$$

Let $u\colon[0,T]\times\mathbb{R}\to\mathbb{R}$ be a function of class $C_b^{1,2}$. Then

$$\begin{aligned}u(t,X_t)=u(0,x)+\int_0^t\left[\frac{\partial u}{\partial s}(s,X_s)+\frac{\partial u}{\partial x}(s,X_s)f(s)+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}(s,X_s)|g(s)|^2\right]ds\\ +\int_0^t\frac{\partial u}{\partial x}(s,X_s)g(s)d\beta(s),\quad (7.19)\end{aligned}$$

for all $t\in[0,T]$.

Proof of Proposition 7.6. We do the proof in the case where u is independent on t and $f\equiv 0$ since the more delicate (and remarkable) term in (7.19) is the Itô's correction involving the second derivative of u. By approximation, it is also sufficient to consider the case where u is in C^3_b and g is the elementary process

$$g=\sum_{l=0}^{m-1} g_l \mathbf{1}_{(s_l,s_{l+1}]},$$

where $(s_l)_{0,m}$ is a subdivision of $[0,T]$ and g_l is a.s. bounded: $|g_l|\leq M$ a.s. Let $\sigma=(t_i)_{0,n}$ be a subdivision of $[0,T]$ which is a refinement of (s_l). Let us consider the case $t=T$ only (for general times t, replace t_i by $t_i\wedge t$ in the formulas below). We decompose

$$u(X_T)-u(x)=\sum_{i=0}^{n-1} u(X_{t_{i+1}})-u(X_{t_i}),$$

and use the Taylor formula to get

$$u(X_T)-u(x)=\sum_{i=0}^{n-1} u'(X_{t_i})(X_{t_{i+1}}-X_{t_i})+\frac{1}{2}u''(X_{t_i})(X_{t_{i+1}}-X_{t_i})^2+r^1_\sigma, \tag{7.20}$$

where

$$|r^1_\sigma|\leq \frac{1}{6}\|u^{(3)}\|_{C_b(\mathbb{R})}\sum_{i=0}^{n-1}|X_{t_{i+1}}-X_{t_i}|^3. \tag{7.21}$$

Since $X_{t_{i+1}}-X_{t_i}=g(t_i)\delta\beta(t_i)$, $\delta\beta(t_i):=\beta(t_{i+1})-\beta(t_i)$, we deduce from (7.20)-(7.21) that

$$u(X_T)-u(x)=\sum_{i=0}^{n-1} u'(X_{t_i})g(t_i)\delta\beta(t_i)+\frac{1}{2}u''(X_{t_i})|g(t_i)|^2|\delta\beta(t_i)|^2+r^1_\sigma, \tag{7.22}$$

and that

$$\mathbb{E}|r^1_\sigma|\leq \frac{1}{6}\|u^{(3)}\|_{C(\mathbb{R})}M\sum_{i=0}^{n-1}\mathbb{E}|\delta\beta(t_i)|^3=\mathcal{O}\left(\sum_{i=0}^{n-1}(t_{i+1}-t_i)^{3/2}\right)=\mathcal{O}(|\sigma|^{1/2}). \tag{7.23}$$

By (7.22), we get

$$u(X_T)-u(x)=\int_0^T u'(X_t)g(t)d\beta(t)+\int_0^T\frac{1}{2}u''(X_t)|g(t)|^2dt+r^3_\sigma+r^2_\sigma+r^1_\sigma, \tag{7.24}$$

where the remainder r^3_σ and r^2_σ are such that

$$\sum_{i=0}^{n-1} u'(X_{t_i})g(t_i)\delta\beta(t_i)=\int_0^T u'(X_t)g(t)d\beta(t)+r^3_\sigma,$$

and

$$\sum_{i=0}^{n-1}\frac{1}{2}u''(X_{t_i})|g(t_i)|^2|\delta\beta(t_i)|^2=\int_0^T\frac{1}{2}u''(X_t)|g(t)|^2dt+r^2_\sigma. \tag{7.25}$$

By Itô's Isometry, we have the estimate

$$\begin{aligned}\mathbb{E}|r^3_\sigma|^2 &= \sum_{i=0}^{n-1}\mathbb{E}\int_{t_i}^{t_{i+1}}|u'(X_t)-u'(X_{t_i})|^2|g(t_i)|^2dt\\&\leq M^2\|u''\|^2_{C_b(\mathbb{R})}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}|X_t-X_{t_i}|^2dt.\end{aligned}$$

Since $\mathbb{E}|X_t-X_{t_i}|^2=\mathbb{E}|g(t_i)|^2(t-t_i)\leq M^2(t-t_i)$, we deduce that

$$\mathbb{E}|r^3_\sigma|^2\leq M^4\|u''\|^2_{C_b(\mathbb{R})}\sum_{i=0}^{n-1}(t_{i+1}-t_i)^2=\mathcal{O}(|\sigma|). \tag{7.26}$$

Some similar estimates show that we can replace X_t by the step function equal to X_{t_i} on $(t_i,t_{i+1}]$ in the right-hand side of (7.25) and that this contributes to an error of order $|\sigma|$: $r^2_\sigma=r^4_\sigma+r^5_\sigma$, where $\mathbb{E}|r^4_\sigma|^2=\mathcal{O}(|\sigma|)$, where the remainder term r^5_σ is defined by

$$r^5_\sigma=\sum_{i=0}^{n-1}\frac{1}{2}u''(X_{t_i})|g(t_i)|^2[|\delta\beta(t_i)|^2-(t_{i+1}-t_i)].$$

Since $(t_{i+1}-t_i)=\mathbb{E}[|\delta\beta(t_i)|^2|\mathcal{F}_{t_i}]$, cancellations occur when we develop the square of r^5_σ and take the expectation: only the pure squares remain, and we get

$$\mathbb{E}|r^5_\sigma|^2\leq\|u''\|^2_{C_b(\mathbb{R})}M^4\sum_{i=0}^{n-1}\mathbb{E}||\delta\beta(t_i)|^2-(t_{i+1}-t_i)|^2=\mathcal{O}(|\sigma|). \tag{7.27}$$

Using (7.23), (7.26), (7.27), we can pass to the limit $|\sigma|\to 0$ in (7.24) to get (7.19) in our simplified case. □

7.5 Generalization in infinite dimension

We have defined the stochastic integral of an Hilbert-valued integrand against a one-dimensional Wiener process. In this section we explain briefly how to generalize this construction and the Itô Formula to higher dimension.

7.5.1 Finite dimension

Let $d\geq 1$. A d-dimensional Wiener process $(B(t))_{t\geq 0}$ admits the decomposition

$$B(t)=\sum_{k=1}^{d}\beta_k(t)e_k, \tag{7.28}$$

where (e_k) is the canonical basis of $\mathbb{R}^d$ and $\beta_1(t),\dots,\beta_d(t)$ are independent one-dimensional processes. Let $(\mathcal{F}_t)_{t\geq 0}$ be a given filtration, such that, for all k, $(\beta_k(t))$ is

$(\mathcal{F}_t)$-adapted, and the increment $\beta_k(t) - \beta_k(s)$ is independent on $\mathcal{F}_s$ for all $0 \leq s \leq t$. Let K be a separable Hilbert space. Let $(g(t))$ be a process with values in $\mathcal{L}(\mathbb{R}^d; K)$ such that

$$g \in L^2(\Omega \times [0,T], \mathcal{P}_T^*; \mathcal{L}(\mathbb{R}^d; K)).$$

We set

$$\int_0^T g(t)dB(t) = \sum_{k=1}^d \int_0^T g(t)e_k d\beta_k(t). \tag{7.29}$$

This defines an element of $L^2(\Omega; K)$ and, using the independence of $\beta_1(t), \dots, \beta_d(t)$, we have the Itô isometry

$$\mathbb{E}\left\|\int_0^T g(t)dB(t)\right\|_K^2 = \sum_{k=1}^d \int_0^T \mathbb{E}\|g(t)e_k\|_K^2 dt. \tag{7.30}$$

Let us examine the generalization of the Itô Formula. We refer to the proof of Proposition 7.6. If $u \in C_b^3(K;\mathbb{R})$, we have the Taylor expansion (which generalizes (7.20))

$$u(X_{t_{i+1}})-u(X_{t_i}) = Du(X_{t_i})\cdot(X_{t_{i+1}}-X_{t_i})+\frac{1}{2}D^2u(X_{t_i})\cdot(X_{t_{i+1}}-X_{t_i})^{\otimes 2}+\mathcal{O}(|X_{t_{i+1}}-X_{t_i}|^3).$$

The increment being here $X_{t_{i+1}} - X_{t_i} = \sum_{1\leq k\leq d} g(t_i)e_k\delta\beta_k(t_i)$, we have to examine in particular the term

$$\sum_{1\leq k,l\leq d} D^2u(X_{t_i}) \cdot (g(t_i)e_k, g(t_i)e_l)\delta\beta_k(t_i)\delta\beta_l(t_i). \tag{7.31}$$

Ii is treated like the left-hand side of (7.25), with the additional fact that the independence of $\beta_1(t), \dots, \beta_d(t)$ comes into play and that the off-diagonal terms in (7.27), the sum over $k \neq l$, is negligible when $|\sigma| \to 0$. We obtain the Itô Formula

$$\begin{aligned} u(t,X_t) = u(0,x) &+ \int_0^t \left[\frac{\partial u}{\partial s}(s,X_s) + Du(s,X_s)\cdot f(s)\right] ds \\ &\sum_{k=1}^d \frac{1}{2}\int_0^t D^2u(s,X_s)\cdot(g(s)e_k, g(s)e_k)ds + \int_0^t Du(s,X_s)\cdot g(s)dB(s), \end{aligned} \tag{7.32}$$

for

$$X_t = x + \int_0^t f(s)ds + \int_0^t g(s)dB(s), \tag{7.33}$$

where D in (7.32) means D_x. In (7.32), $u\colon [0,T]\times K \to \mathbb{R}$ is of class $C_b^{1,2}$. In (7.32) and (7.33), the integrands are in the following classes:

$$f \in L^1(\Omega\times[0,T], \mathcal{P}_T^*; K), \quad g \in L^2(\Omega\times[0,T], \mathcal{P}_T^*; \mathcal{L}(\mathbb{R}^d; K))).$$

A standard instance of (7.32) and (7.33) is when K is finite dimensional, $K = \mathbb{R}^m$ (often with $m = d$). Then $g(t) \in \mathcal{L}(\mathbb{R}^d;\mathbb{R}^m)$ is assimilated with its matrix representation ($d\times m$

matrix) in the canonical bases of $\mathbb{R}^d$ and $\mathbb{R}^m$, $D^2u(t,x)$, which is a bilinear form on $\mathbb{R}^m$ is assimilated to a $m\times m$ matrix, and the Itô correction term rewritten

$$\sum_{k=1}^{d}\frac{1}{2}D^2u(s,X_s)\cdot(g(s)e_k,g(s)e_k)=\frac{1}{2}\mathrm{Trace}(g(s)^*D^2u(s,X_s)g(s)). \tag{7.34}$$

7.5.2 Infinite dimension

Cylindrical Wiener process Let H be a separable Hilbert space with an orthonormal basis $(e_k)_{k\geq 1}$. Let U be an other Hilbert space such that $H\hookrightarrow U$ with Hilbert-Schmidt injection. Recall (see [Bre11, p. 497]) that an operator $\Phi\colon H\to K$ (K is an other Hilbert space here) is said to be Hilbert-Schmidt if

$$\sum_{k\geq 1}\|\Phi e_k\|_K^2<+\infty. \tag{7.35}$$

We denote by $\mathcal{L}_2(H;K)$ the class of Hilbert-Schmidt operators from H to K. This is a Hilbert space[9] for the scalar product

$$\langle\Phi,\Psi\rangle_{\mathcal{L}_2(H;K)}=\sum_{k\geq 1}\langle\Phi e_k,\Psi e_k\rangle_K. \tag{7.36}$$

The scalar product, and thus the norm $\|\Phi\|^2_{\mathcal{L}_2(H;K)}$ in (7.35), is independent on the choice of the orthonormal basis $(e_k)_{k\geq 1}$ on H (see [Bre11, p. 497] again). Let $(\beta_k(t))_{k\geq 1}$ be independent one-dimensional Wiener processes. If $\Phi\in\mathcal{L}_2(H;K)$, we set

$$W(t)=\sum_{k\geq 1}\beta_k(t)e_k,\quad \Phi W(t)=\sum_{k\geq 1}\beta_k(t)\Phi e_k. \tag{7.37}$$

A formal computation, using independence, gives, for $t>0$,

$$\mathbb{E}\|W(t)\|_H^2=\sum_{k\geq 1}\mathbb{E}|\beta_k(t)|^2\|e_k\|_H^2=\sum_{k\geq 1}t=+\infty.$$

Therefore, the process $W(t)$ is not well defined in H. However, $\Phi W(t)$ is well defined as a process over K. More exactly, it is well defined in the space $L^2(\Omega;K)$ since, using independence,

$$\mathbb{E}\left\|\sum_{p\leq k\leq q}\beta_k(t)\Phi e_k\right\|_K^2=t\sum_{p\leq k\leq q}\|\Phi e_k\|_K^2,$$

which gives a Cauchy condition in $L^2(\Omega;K)$ for the series defining $\Phi W(t)$. Taking $p=1$ and sending q to $+\infty$, we see also that

$$\mathbb{E}\|\Phi W(t)\|_K^2=t\|\Phi\|^2_{\mathcal{L}_2(H;K)}.$$

Since the injection $H\hookrightarrow U$ is Hilbert-Schmidt by hypothesis, $W(t)$ is well defined in $L^2(\Omega;U)$. We call $W(t)$ a *cylindrical Wiener process.*

[9] a separable Hilbert space, with orthonormal basis $(\Phi_{kl})_{k,l\geq 1}=(e_k\otimes f_l)_{k,l\geq 1}$, $\Phi_{k,l}u=\langle u,e_k\rangle_H f_l$, where $(f_l)_{l\geq 1}$ is an orthonormal basis of K

Stochastic integral We generalize (7.29) into

$$\int_0^T g(t)dW(t) = \sum_{k\geq 1}\int_0^T g(t)e_k d\beta_k(t), \tag{7.38}$$

where $g \in L^2(\Omega \times [0,T], \mathcal{P}_T^*; \mathcal{L}_2(H;K))$. This defines an element of $L^2(\Omega; K)$. We have the Itô isometry

$$\mathbb{E}\left\|\int_0^T g(t)dW(t)\right\|_K^2 = \sum_{k\geq 1}\int_0^T \mathbb{E}\|g(t)e_k\|_K^2 dt = \int_0^T \mathbb{E}\|g(t)\|_{\mathcal{L}_2(H;K)}^2 dt. \tag{7.39}$$

The generalization of (7.32) is

$$\begin{aligned} u(t,X_t) = u(0,x) + \int_0^t \left[\frac{\partial u}{\partial s}(s,X_s) + Du(s,X_s)\cdot f(s)\right] ds \\ \sum_{k\geq 1}\frac{1}{2}\int_0^t D^2 u(s,X_s)\cdot(g(s)e_k, g(s)e_k)ds + \int_0^t Du(s,X_s)\cdot g(s)dW(s), \end{aligned} \tag{7.40}$$

for

$$X_t = x + \int_0^t f(s)ds + \int_0^t g(s)dW(s), \tag{7.41}$$

where $u \in C_b^{1,2}([0,T]\times K;\mathbb{R})$. Note that the stochastic integral in (7.40) makes sense since $\mathcal{L}_2(H;K)$ is a left ideal: $Du(s,X_s) \in \mathcal{L}(K)$ and $D^2u(s,X_s)\cdot g(s) \in \mathcal{L}_2(H;K)$. Note also that we have a formula analogous to (7.34):

$$\sum_{k\geq 1}\frac{1}{2}D^2u(s,X_s)\cdot(g(s)e_k, g(s)e_k) = \frac{1}{2}\mathrm{Trace}(g(s)^* D^2u(s,X_s)g(s)), \tag{7.42}$$

where an operator T in $\mathcal{L}(K;K)$ is said to be trace-class if, for a given orthonormal basis (e_k') of K, the series of general term $\langle Te_k', e_k'\rangle$ is absolutely convergent, with the definition

$$\mathrm{Trace}(T) = \sum_{k\geq 1}\langle Te_k', e_k'\rangle.$$

8 Stochastic differential equations

Let H, K be some separable Hilbert spaces, let $(W(t))$ be a cylindrical Wiener process as in (7.37). For some general integrands

$$f \in L^1(\Omega\times[0,T];K), \quad g \in L^2(\Omega\times[0,T], \mathcal{P}_T^*; \mathcal{L}_2(H;K))),$$

we use the differential notation

$$dX_t = f(t)dt + g(t)dW(t) \tag{8.1}$$

to mean that (7.41) is satisfied for all $t \in [0,T]$, for a given $x \in K$. The meaning of

$$dX_t = f(X_t)dt + g(X_t)dW(t),\ t \in [0,T] \tag{8.2}$$
$$X_0 = x, \tag{8.3}$$

is therefore

$$X_t = x + \int_0^t f(X_s)ds + \int_0^t g(X_s)dW(s), \tag{8.4}$$

for all $t \in [0,T]$. We will study the Cauchy Problem (8.2)-(8.3) in the case where x is a given $\mathcal{F}_0$-measurable random variable. Since x rather stands for an arbitrary point in $\mathbb{R}^d$. We denote by

$$X_0 = \xi \tag{8.5}$$

the Cauchy condition, and consider (8.2)-(8.5) under the integral form

$$X_t = \xi + \int_0^t f(X_s)ds + \int_0^t g(X_s)dW(s), \tag{8.6}$$

8.1 Resolution

We will solve (8.2) first in the case where f and g are two Lipschitz continuous functions. With some applications and the theory of ordinary differential equations in mind, it would be natural, then, to study the case of locally Lipschitz functions f and g. We refer to[10] on that subject. We are more interested in the case where f is a (linear) differential operator. In a second time therefore, we will consider the case where f is the sum of an unbounded operator and of a Lipschitz continuous function, g being a Lipschitz continuous function.

8.2 The global Lipschitz case

Definition 8.1 (Solution to the Cauchy Problem)**.** Let $f\colon K \to K$, $g\colon K \to \mathcal{L}_2(H;K)$ be some Lipschitz continuous functions. Let $\xi \in L^2(\Omega;\mathcal{F}_0)$. An adapted process $X \in C([0,T];L^2(\Omega;K))$ is said to be solution to (8.2)-(8.5) if, for all $t \in [0,T]$, (8.6) is satisfied a.s.

If $X \in C([0,T];L^2(\Omega;K))$ is adapted, then

$$f(X) \in C([0,T];L^2(\Omega;K)), \quad g(X) \in C([0,T];L^2(\Omega;\mathcal{L}_2(H;K)))$$

are adapted. In particular, they are admissible integrands in (8.4). They define some adapted processes which are also in the class $C([0,T];L^2(\Omega;K))$. Indeed, assuming

$$f(0) = 0, \quad g(0) = 0, \tag{8.7}$$

[10] TODO ref.

we have, for $0 \leq s \leq t \leq T$,

$$\mathbb{E}\left\|\int_s^t f(X_\sigma)d\sigma\right\|_K^2 \leq (t-s)\int_s^t \mathbb{E}\|f(X_\sigma)\|_K^2 d\sigma \leq (t-s)^2 \mathrm{Lip}(f)^2 \sup_{\sigma\in[0,T]} \mathbb{E}\|X_\sigma\|_K^2, \quad (8.8)$$

and

$$\mathbb{E}\left\|\int_s^t g(X_\sigma)dW(\sigma)\right\|_K^2 = \int_s^t \mathbb{E}\|g(X_\sigma)\|_{\mathcal{L}_2(H;K)}^2 d\sigma \leq (t-s)\mathrm{Lip}(g)^2 \sup_{\sigma\in[0,T]} \mathbb{E}\|X_\sigma\|_K^2. \quad (8.9)$$

If (8.7) is not satisfied, the estimate (8.8) holds true if we replace $\mathrm{Lip}(f)^2$ by $2\mathrm{Lip}(f)^2 + 2\|f(0)\|_K^2$, and similarly for (8.9) .

Theorem 8.1 (The Cauchy problem for SDEs, global Lipschitz case)**.** *Let H, K be some separable Hilbert spaces, let $(W(t))$ be a cylindrical Wiener process as in (7.37). Let $f\colon K \to K$, $g\colon K \to \mathcal{L}_2(H;K)$ be some Lipschitz continuous functions. Let $\xi \in L^2(\Omega;\mathcal{F}_0)$. Then the Cauchy Problem (8.2)-(8.5) has a unique solution*

$$X \in C([0,T];L^2(\Omega;K)).$$

Two solutions X and $\tilde{X}$ issued from two data ξ and $\tilde{\xi}$ satisfy the estimate

$$\mathbb{E}\|X(t)-\tilde{X}(t)\|_K^2 \leq C\mathbb{E}\|\xi-\tilde{\xi}\|_K^2, \quad (8.10)$$

for all $t \in [0,T]$, where C is a constant depending on T, $\mathrm{Lip}(f)$, $\mathrm{Lip}(g)$.

Proof of Theorem 8.1. Let $\mathcal{T}(X)_t$ denote the right-hand side of (8.6). By (8.8), (8.9), this defines an element of the Banach space E constituted of the adapted functions $X \in C([0,T];L^2(\Omega;K))$. Let us consider the norm, for $M > 0$,

$$\|X\|_E = \sup_{t\in[0,T]} e^{-Mt}\left(\mathbb{E}\|X_t\|_K^2\right)^{1/2}$$

on E. Instead of (8.8), we write

$$\begin{aligned}\mathbb{E}\left\|\int_0^t [f(X_\sigma)-f(\tilde{X}_\sigma)]d\sigma\right\|_K^2 &\leq t\mathrm{Lip}(f)^2\int_0^t \mathbb{E}\|X_\sigma-\tilde{X}_\sigma\|_K^2 d\sigma \\ &\leq \frac{t(e^{2Mt}-1)}{2M}\mathrm{Lip}(f)^2\|X-\tilde{X}\|_E^2 \\ &\leq \frac{T}{2M}\mathrm{Lip}(f)^2\|X-\tilde{X}\|_E^2.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathbb{E}\left\|\int_0^t [g(X_\sigma)-g(\tilde{X}_\sigma)]dW(\sigma)\right\|_K^2 &\leq \mathrm{Lip}(g)^2\int_0^t \mathbb{E}\|X_\sigma-\tilde{X}_\sigma\|_K^2 d\sigma \\ &\leq \frac{(e^{2Mt}-1)}{2M}\mathrm{Lip}(g)^2\|X-\tilde{X}\|_E^2.\end{aligned}$$

We multiply the previous estimates by e^{-2Mt} and take the sup over $t \in [0,T]$. It follows that $\mathcal{T}$ is a k-contraction on E, with

$$k = M^{-1} \max(1, \sqrt{T}) \sqrt{\mathrm{Lip}(f)^2 + \mathrm{Lip}(g)^2}.$$

For M large enough, $k < 1$ and the Banach fixed-point theorem gives the result. The estimate (8.10) is obtained by the Grönwall Lemma. □

Remark 8.1 (Iteration). By construction, the solution $(X_t(\xi))$ to (8.2)-(8.5) is the limit in $C([0,T]; L^2(\Omega; K))$ of the iterative sequence $(X_t^n(\xi))$ defined by $X_t^{n+1}(\xi) = \mathcal{T}(X_t^n(\xi))$, $X_t^0(\xi) = \xi$.

Exercise 8.2. Compute the solutions to the following SDEs

1. Ornstein-Uhlenbeck:
$$dX_t = -X_t + \sqrt{2} dB_t,$$
with initial condition $X_0 = x \in \mathbb{R}^d$ (*Hint:* use Duhamel's integral formula). Show that X_t is a Gaussian variable, and give the parameters (*Note:* you may use the result of Question 3. of Exercise 7.1). Find the limit in law of X_t when $[t \to +\infty]$.

2. the equation
$$dX_t = \sigma X_t d\beta_t,$$
with initial condition $X_0 = x \in \mathbb{R}$, where $\sigma > 0$ (*Hint:* apply Itô's Formula to $\ln(|X_t|)$).

The solution to Exercise 8.2 is here.

8.3 Markov property, generator

Theorem 8.2. *Under the hypotheses of Theorem 8.1, let $X_t(\xi)$ be the unique solution to the Cauchy Problem (8.2)-(8.5). Then $(X_t(\xi))$ is a Markov process relatively to $(\mathcal{F}_t)$ with transition semi-group (Π_t) given by $\Pi_t \varphi(x) = \mathbb{E}\varphi(X_t(x))$ for all $\varphi \in \mathrm{BM}(K)$.*

Proof of Theorem 8.2. Let us denote by $X(t, s; \xi)$ the value at time t of the solution to (8.2) issued from ξ at time s ($\xi \in L^2(\Omega; \mathcal{F}_s)$):

$$X(t, s; \xi) = \xi + \int_s^t f(X(\sigma, s; \xi)) d\sigma + \int_s^t g(X(\sigma, s; \xi)) dW(\sigma).$$

By uniqueness, we have the semi-group property

$$X(t + s, 0; x) = X(t + s, s; y), \quad y = X(s, 0; x). \tag{8.11}$$

Our aim is to prove that, for $0 \le s, t$

$$\mathbb{E}\left[\varphi(X(t + s, 0; x)) | \mathcal{F}_s\right] = (\Pi_t \varphi)(X(s, 0; x)), \tag{8.12}$$

where $\varphi \in \mathrm{BM}(K)$ is given. Consider first the case $\varphi \in C_b(K)$. Using the semi-group property (8.11), the Markov property (8.12) amounts to

$$\mathbb{E}\left[\varphi(X(t+s,s;\xi))|\mathcal{F}_s\right] = (\Pi_t\varphi)(\xi), \tag{8.13}$$

in the special case $\xi = X(s,0;x)$. We will show that (8.13) is actually true for all square integrable, $\sigma(X(s,0;x))$-measurable random variable ξ. By (8.10), Π_t is Feller: $\Pi_t\colon C_b(K) \to C_b(K)$. Indeed, if $\psi \in C_b(K)$, then $\psi(X_t(y)) \to \psi(X_t(x))$ a.s. if $y \to x$ in $\mathbb{R}^d$ (we use (8.10) and the continuity of ψ). Since ψ is bounded, we deduce that $\mathbb{E}\psi(X_t(y)) \to \mathbb{E}\psi(X_t(x))$ by dominated convergence. Consequently, it is sufficient to establish (8.13) for a dense subset of $L^2(\Omega;\sigma(X(s,0;x)))$. We consider the set of simple functions. Let

$$\xi = \sum_{k=1}^n x_k \mathbf{1}_{A_k}, \quad x_k \in \mathbb{R}^d, \quad A_k \in \sigma(X(s,0;x)).$$

Then

$$X(t+s,s;\xi) = \sum_{k=1}^n X(t+s,s;x_k)\mathbf{1}_{A_k} \quad \text{a.s.} \tag{8.14}$$

To prove (8.14), we can assume $s=0$ for simplicity. Remind that $X(t,0;\xi)$ is the limit of the iterative sequence $(X_t^n(\xi))$, defined by $X_t^{n+1}(\xi) = \mathcal{T}(X_t^n(\xi))$, $X_t^0(\xi) = \xi$ (see Remark 8.1). It is sufficient therefore to check (8.14) on the members of the sequence $(X_t^n(\xi))$, and this is not difficult, using recursion on n. By (8.14), we have

$$\mathbb{E}\left[\varphi(X(t+s,s;\xi))|\mathcal{F}_s\right] = \sum_{k=1}^n \mathbb{E}\left[\varphi(X(t+s,s;x_k))\right]\mathbf{1}_{A_k}. \tag{8.15}$$

Admit that

$$\mathbb{E}\left[\varphi(X(t+s,s;x))\right] = \mathbb{E}\left[\varphi(X(t,0;x))\right] = \Pi_t\varphi(x). \tag{8.16}$$

Then (8.13) follows from (8.15). The basic reason for (8.16) is that $(\beta(t+s))_{t\geq 0}$ and $(\beta(t))_{t\geq 0}$ have the same increments. To prove (8.16), we can once again consider the iterative sequence (X^n). We want to prove that $\mathrm{Law}(X^n(t+s,s;x)) = \mathrm{Law}(X^n(t,0;x))$. We establish this identity by recursion on n. It is true for $n=0$ since both random variables are equal to x then. If this is true at rank n, then we consider $X^{n+1}(t+s,s;x)$. It is given as

$$X^{n+1}(t+s,s;x) = x + \int_s^{t+s} f(X^n(r,s;x))dr + \int_s^{t+s} g(X^n(r,s;x))d\beta(r).$$

Since $r \mapsto X^n(r,s;x) \in C([0,T];L^2(\Omega;K))$, $X^{n+1}(t+s,s;x)$ is the limit when $|\sigma| \to 0$ of

$$\begin{aligned} X_\sigma^{n+1}(t+s,s;x) := x &+ \sum_{i=0}^{m-1}(t_{i+1}-t_i)f(X^n(t_i+s,s;x)) \\ &+ \sum_{i=0}^{m-1}(\beta(t_{i+1}+s)-\beta(t_i+s))g(X^n(t_i+s,s;x)). \end{aligned}$$

where $\sigma = (t_i)_{0,m}$ is a subdivision of $[0,t]$. We have

$$(\beta(t_{i+1}+s)-\beta(t_i+s), X^n(t_i+s,s;x)) \overset{\text{Law}}{=} (\beta(t_{i+1})-\beta(t_i), X^n(t_i,0;x))$$

because the components are independent and identical in law. We deduce, with obvious notations, that

$$X^{n+1}_\sigma(t+s,s;x) \overset{\text{Law}}{=} X^{n+1}_\sigma(t,0;x).$$

At the limit $|\sigma|\to 0$, we obtain the desired result. Consequently, (8.12) is established for $\varphi \in C_b(E)$. The end of the proof is as in Section 6.2.2 and follows four steps:

1. like in Lemma 6.4, we prove first that (Π_t) satisfies the points 1, 2, 3 of Definition 4.5 (this uses only the definition $\Pi_t\varphi(x) = \mathbb{E}\varphi(X(t,0;x))$),
2. then we use (8.12) and an argument of separating class to obtain the semi-group property for (Π_t),
3. Proposition 4.3 shows that (Π_t) is a Markov semi-group,
4. we deduce (8.12) for general $\varphi \in \mathrm{BM}(\mathbb{R}^d)$ by an argument of separating class .

This concludes the proof of Theorem 8.2. □

Remark 8.2 (Stochastic continuity). Since $(X_t) \in C([0,T];L^2(\Omega;K))$, $t\mapsto \Pi_t\varphi(x)$ is continuous on $\mathbb{R}_+$ for all $\varphi\in C_b(E)$.

Proposition 8.3. *Under the hypotheses of Theorem 8.1, let $(X_t(x))$ be the unique solution to the Cauchy Problem (8.2)-(8.3), (Π_t) its transition semi-group. Let $\mathcal{L}$ be the generator associated to (Π_t). Let $\mathcal{L}_0$ be the unbounded operator defined by its domain*

$$\mathcal{D}(\mathcal{L}_0) = \left\{\varphi\in C^2_b(K); \sup_{x\in K}\left[\|x\|_K\|D\varphi(x)\|_{\mathcal{L}(K;K)} + \|x\|_K^2\|D^2\varphi(x)\|_{\mathcal{L}(K\times K;K)}\right] < +\infty\right\}, \tag{8.17}$$

and its value

$$\mathcal{L}_0\varphi(x) = D\varphi(x)\cdot f(x) + \frac{1}{2}\mathrm{Trace}(g(x)^* D^2\varphi(x) g(x)), \tag{8.18}$$

for $\varphi\in\mathcal{D}(\mathcal{L}_0)$. Then $\mathcal{L}\supset\mathcal{L}_0$.

Proof of Proposition 8.3. Let $\varphi\in\mathcal{D}(\mathcal{L}_0)$. We will prove that

$$\Pi_t\varphi(x) = \varphi(x) + t\mathcal{L}_0\varphi(x) + t\eta_t(x), \tag{8.19}$$

where (η_t) is π-converging when $t\to 0$ on K. We apply the Itô Formula (7.40), using the notation (7.42). Taking the expectation, we obtain

$$\begin{aligned}\Pi_t\varphi(x) &= \varphi(x) + \int_0^t \mathbb{E}\left[\mathcal{L}\varphi(X_s(x))\right] ds\\ &= \varphi(x) + t\mathcal{L}\varphi(x) + t\eta_t(x), \quad \eta_t(x) := \int_0^1 \mathbb{E}\left[\mathcal{L}\varphi(X_{st}(x)) - \mathcal{L}\varphi(x)\right] ds.\end{aligned}$$

Since $\varphi\in\mathcal{D}(\mathcal{L}_0)$, the function $\mathcal{L}\varphi$ is bounded, continuous. Since $X_{st}(x)\to x$ almost surely, and for almost all $s\in[0,1]$ when $t\to 0$, we obtain $\eta_t \xrightarrow{\pi} 0$ by dominated convergence. □

A Maximal coupling

Let (F, d) be a metric space which is separable and complete. If μ and ν are two Borel probability measures on F, the total variation of the signed measure $\mu - \nu$ is

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_A |\mu(A) - \nu(A)|, \tag{A.1}$$

where the sup is taken over Borel subsets A of F. If X, Y are two random variables with respective law μ and ν, then

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \mathbb{P}(X \neq Y). \tag{A.2}$$

Indeed, if A is a Borel set in F, then

$$\begin{aligned}\mu(A) = \mathbb{P}(X \in A) &= \mathbb{P}(X \in A \text{ and } X = Y) + \mathbb{P}(X \in A \text{ and } X \neq Y) \\ &\leq \mathbb{P}(Y \in A) + \mathbb{P}(X \neq Y) = \nu(A) + \mathbb{P}(X \neq Y).\end{aligned}$$

If the equality in (A.2) is realized, then (X, Y) is said to be a *maximal coupling* of (μ, ν).

Example Let μ be the uniform measure on $[0, 1]$, ν the uniform measure on $[0, 1/2]$. What is $\|\mu - \nu\|_{\mathrm{TV}}$? Let Y be a random variable of law ν and B an independent Bernoulli random variable: $\mathbb{P}(B = \pm 1) = \frac{1}{2}$. With Y and B, construct a maximal coupling (X, Y) of (μ, ν) (*cf.* Exercise 2.15).

Theorem A.1 (Dobrushin's maximal coupling theorem). *There exists a maximal coupling (X, Y) of (μ, ν).*

Proof of Theorem A.1. Let $\lambda = \mu + \nu$. Then μ and ν are absolutely continuous with respect to λ. By the Radon-Nikodym theorem, μ and ν admits some densities f and g, respectively, with respect to λ. Since all the measures are positive, and since $\mu(F) = \nu(F) = 1$, we have

$$f, g \geq 0 \ \lambda - \text{a.e.}, \qquad \int_F f d\lambda = \int_F g d\lambda = 1. \tag{A.3}$$

Let $A = \{f \geq g\}$ and $B = \{f < g\}$. By definition of the total variation distance (A.1), we have

$$\|\mu - \nu\|_{\mathrm{TV}} = \max\left[\int_A (f - g) d\lambda, \int_B (g - f) d\lambda\right].$$

By the normalization condition (A.3), the two quantities in the max are equal. Therefore

$$\|\mu - \nu\|_{\mathrm{TV}} = \int_A (f - g) d\lambda = \int_B (g - f) d\lambda. \tag{A.4}$$

Using the formula $(f - g)^+ = f - f \wedge g$ and the normalization property (A.3) on f, we have also the equation

$$\|\mu - \nu\|_{\mathrm{TV}} = \int_A (f - g) d\lambda = \int_F (f - g)^+ d\lambda = 1 - \kappa, \qquad \kappa := \int_F f \wedge g d\lambda. \tag{A.5}$$

If $\kappa = 0$, then μ and ν are mutually singular and any coupling is a maximal coupling. We consider the non-trivial case $\kappa > 0$. Let U, η, ξ, ζ be some independent random variables with the following laws: U has the uniform law on $[0, 1]$,

$$\eta \sim \frac{1}{\kappa} f \wedge \wedge g, \quad \xi \sim \frac{f-g}{1-\kappa} \mathbf{1}_A = \frac{f - f \wedge g}{1-\kappa}, \quad \zeta \sim \frac{g-f}{1-\kappa} \mathbf{1}_B = \frac{g - f \wedge g}{1-\kappa}.$$

Draw U, η, ξ, ζ. If $U \leq \kappa$, set $X = Y = \eta$. Otherwise, set $X = \xi$, $Y = \zeta$. Then, if D is a Borel subset of F, we have

$$\begin{aligned}
\mathbb{P}(X \in D) &= \mathbb{P}(X \in D | U \leq \kappa) \mathbb{P}(U \leq \kappa) + \mathbb{P}(x \in D | U > \kappa) \mathbb{P}(U > \kappa) \\
&= \kappa \mathbb{P}(\eta \in D) + (1 - \kappa) \mathbb{P}(\xi \in D) \\
&= \int_D f \wedge g d\lambda + \int_D (f - f \wedge g) d\lambda = \int_D f d\lambda = \mu(A).
\end{aligned}$$

Similarly, we show that Y has law ν. Since A and B are disjoint, $X = Y$ if, and only if, $U \leq \kappa$ and thus

$$\|\mu - \nu\|_{\mathrm{TV}} = 1 - \kappa = \mathbb{P}(X \neq Y).$$

$\square$

B Solution to the exercises

Solution to Exercise 2.2.

1. $\Omega = \{1, \dots, 6\}$, $\mathbb{P}(\{i\}) = \frac{1}{6}$, $A = \{2, 4, 6\}$. The experiment is *rolling a dice*, A is the event *"the outcome is an even number"*.

2. $\Omega = \{H, T\}^2$, $\mathbb{P}(\{\omega\}) = \frac{1}{4}$ for each $\omega \in \Omega$, $A = \{(H, T), (H, H)\}$. The experiment is *tossing two times a unbiased coin* (T stands for "tail" then, and H for "head"). The event A is *"the result of the first tossing is head"*.

3. $\Omega = \{\gamma \in C([0, T]; \mathbb{R}^2); \gamma(0) = 0\}$, $\mathbb{P}$ =to be seen later,

$$A = \{\gamma \in \Omega; \exists t \in [0, T], \gamma(t) \in D\},$$

where D is a closed subset of $\mathbb{R}^2$ (*e.g.* D is the closed disk of radius 1 and center $(2, 0)$). The experiment is *drawing a curve in the plane*. The event A is *"the curve intersects D"*.

To answer to the last question of the exercise about the choice of the σ-algebra $\mathcal{F}$. One natural choice is to consider the Borel σ-algebra. Indeed, endowed with the norm

$$\|\gamma\| = \sup_{t \in [0,T]} |\gamma(t)|,$$

where $|\cdot|$ is the euclidean norm on $\mathbb{R}^2$, the space Ω is a Banach space. The probability measure $\mathbb{P}$ on Ω which we will consider is the Wiener measure. See Section 3.3 on those topics.

Back to Exercise 2.2.

Solution to Exercise 2.4.

1. $\Omega = \{1, \dots, 6\}$, $\mathbb{P}(\{i\}) = \frac{1}{6}$, $A = \{2, 4, 6\}$. Let X =number on the dice. Then $A = \{X \text{ even}\}$.

2. $\Omega = \{H, T\}^2$, $\mathbb{P}(\{\omega\}) = \frac{1}{4}$ for each $\omega \in \Omega$, $A = \{(H, T), (H, H)\}$. Let X ="result of the first tossing". Then $A = \{X = H\}$.

3. $\Omega = \{\gamma \in C([0, T]; \mathbb{R}^2); \gamma(0) = 0\}$, $\mathbb{P}$ =to be seen later,

$$A = \{\gamma \in \Omega; \exists t \in [0, T], \gamma(t) \in D\},$$

where D is a closed subset of $\mathbb{R}^2$ (*e.g.* D is the closed disk of radius 1 and center $(2, 0)$). Let τ (we use the letter τ, more common in that context, instead of X) be defined by

$$\tau = \inf \{t \in [0, T]; \gamma(t) \in D\},$$

with the convention that $\tau = +\infty$ if γ does not intersects D. Note that τ is a random variable if we take for $\mathcal{F}$ the σ-algebra described in the correction of Exercise 2.2 above (*i.e.* the topology of the uniform convergence is considered on Ω). Indeed,

$$\tau = \lim_{n\to+\infty} \inf \{t \in \{t_1, \ldots, t_n\}; \gamma(t) \in D\},$$

where $\{t_i, i \geq 1\}$ is a dense subset of $[0, T]$. The random variable τ is the *hitting time of* D. The event A is now $\{\tau < +\infty\}$.

Back to Exercise 2.4.

Solution to Exercise 2.8.

1. That $\mu_0 = \delta_0$ means that X_0 always take the value 0 (X_0 is deterministic). We have then $X_1 = \pm 1$ with equi-probability, so

$$\mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1},$$

which is an example of Bernoulli's Law $b(\frac{1}{2})$. We have then

$$\mathbb{P}(X_2 = -2) = \frac{1}{4}, \quad \mathbb{P}(X_2 = 0) = \frac{1}{2}, \quad \mathbb{P}(X_2 = +2) = \frac{1}{4}.$$

The law of X_2 is therefore

$$\mu_2 = \frac{1}{4}\left[\delta_{-3/2} + \delta_{-1/2} + \delta_{1/2} + \delta_{3/2}\right].$$

2. The law μ_N is

$$\mu_N = \frac{1}{2^{N+1}}\delta_{-2} + \sum_{-2^{N-1}<k<2^{N-1}} \frac{1}{2^N}\delta_{\frac{k}{2^{N-2}}} + \frac{1}{2^{N+1}}\delta_{-2}. \tag{B.1}$$

3. The answer is that μ_0 is the uniform law on $[-2, 2]$:

$$\mu_0(A) = \frac{1}{4}|A \cap [-2, 2]|,$$

where $|A|$ is the Lebesgue measure of a Lebesgue set $A \subset \mathbb{R}$ (see the proof below for μ_∞). This answer can be simply guessed by examination of the evolution of the process (X_n). An other way to find the right μ_0 is to look at μ_N for large N. Indeed, finding μ_0 such that $\mu_N = \ldots = \mu_1 = \mu_0$ is finding an *equilibrium* to the equation of evolution of (μ_N) (we will not write this latter equation here). Such a μ_0 is called an *invariant measure*. A usual way to find an equilibrium for a system in evolution is to look as the behaviour for large times: if there is convergence to a limit object, this will most probably be an equilibrium of the system. Here, for example, one can look at the evolution starting from the binomial $b(1/2)$ with values in $\{-2,+2\}$, as in Question 2. If $\varphi \in C_b(\mathbb{R})$, then

$$\begin{aligned}\int_{\mathbb{R}} \varphi d\mu_N &= \sum_{-2^{N-1}<k<2^{N-1}} \frac{1}{2^N}\varphi\left(\frac{k}{2^{N-2}}\right) + o(1)\\ &= \frac{1}{4}\sum_{-2^{N-1}<k<2^{N-1}} \frac{1}{2^{N-2}}\varphi\left(\frac{k}{2^{N-2}}\right) + o(1).\end{aligned}$$

We recognize a Riemann sum, which converges to

$$\int_{\mathbb{R}} \varphi d\mu_\infty := \frac{1}{4}\int_{-2}^{2} \varphi(x)dx.$$

The limit law μ_∞ is an invariant measure for good. Indeed, if $X_0 \sim \mu_\infty$, then, by the formula of total probability,

$$\begin{aligned}\mathbb{P}(X_1 \in A) &= \mathbb{P}(X_1 \in A|Z_1 = -1)\mathbb{P}(Z_1 = -1) + \mathbb{P}(X_1 \in A|Z_1 = +1)\mathbb{P}(Z_1 = +1)\\ &= \frac{1}{2}\mathbb{P}(X_1/2 \in A+1) + \frac{1}{2}\mathbb{P}(X_1/2 \in A-1),\end{aligned}$$

for any Borel subsets A of $\mathbb{R}$. This gives

$$8\mathbb{P}(X_1 \in A) = |A_+ \cap [-2,2]| + |A_- \cap [-2,2]|, \quad A_\pm := 2A \pm 2.$$

We compute, thanks to the invariance by translation of the Lebesgue measure and the change of variable formula,

$$|A_+ \cap [-2,2]| = |2A \cap [-4,0]| = 2|A \cap [-2,0], \quad |A_- \cap [-2,2] = 2|A \cap [0,2]|.$$

If follows that $\mathbb{P}(X_1 \in A) = \frac{1}{4}|A \cap [-2,2]| = \mu_\infty(A)$: X_1 has law μ_∞.

Back to Exercise 2.8.

Solution to Exercise 2.9. Any $A \in \sigma(X)$ has the form $X^{-1}(B)$. Hence

$$\mathbb{P}(A) = \mathbb{P}(X^{-1}(B)) = \mu_X(B).$$

Back to Exercise 2.9.

Solution to Exercise 2.11. Since the events A_i form a partition (up to a negligible event) of Ω, the sets $A \cap A_i$ form a partition (up to a negligible event) of A. Therefore $\mathbb{P}(A)$ is the sum of the probabilities $\mathbb{P}(A \cap A_i)$, which are equal to $\mathbb{P}(A|A_i)\mathbb{P}(A_i)$ by definition of the conditional probability. If $\mathbb{P}(A_i) = 0$, then $\mathbb{P}(A|A_i)$ is not defined, but the formula of the total probabilities remain true if we set $\mathbb{P}(A|A_i)\mathbb{P}(A_i) = 0$.

Back to Exercise 2.11.

Solution to Exercise 2.13. We list the outcomes corresponding to A_1 and A_2:

$$A_1 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}.$$

Hence $\mathbb{P}(A_1) = \frac{5}{36}$, $\mathbb{P}(B) = \frac{1}{6}$ and $\mathbb{P}(A_1 \cap B) = \frac{1}{36} \neq \mathbb{P}(A)\mathbb{P}(B)$. In A_2, there are the six elements $(1,6)$, $(2,5)$,... $(6,1)$ and we obtain $\mathbb{P}(A_2) = \frac{1}{6}$, $\mathbb{P}(A_2 \cap B) = \frac{1}{36} = \mathbb{P}(A_2)\mathbb{P}(B)$.
Back to Exercise 2.13.

Solution to Exercise 2.15. Draw $\hat{Y} = Y$. Draw a random variable $Z \in \{-1,+1\}$ of law $b(1/2)$ independently on Y (this corresponds to the tossing of a coin). Set $\hat{X} = Y$ if $Z = +1$ and $\hat{X} = Y + 1/2$ if $Z = -1$. Then $\hat{X}$ has the law of X and $\mathbb{P}(\hat{X} = \hat{Y}) = \frac{1}{2}$. The last assertion is clear, since

$$\mathbb{P}(\hat{X} = \hat{Y}) = \mathbb{P}(Z = +1) = \frac{1}{2}.$$

This is the maximal probability that $X = Y$ since $\{X = Y\} \subset \{X \in [0,1/2]\}$. Let us prove that $\hat{X}$ has the law of X. We use the formula of total probabilities: if A is a Borel subset of $\mathbb{R}$, then

$$\begin{aligned}\mathbb{P}(\hat{X} \in A) &= \mathbb{P}(\hat{X} \in A|Z = +1)\mathbb{P}(Z = +1) + \mathbb{P}(\hat{X} \in A|Z = -1)\mathbb{P}(Z = -1)\\ &= \frac{1}{2}\mathbb{P}(Y \in A) + \frac{1}{2}\mathbb{P}(Y \in A - 1/2).\end{aligned}$$

The first term $\frac{1}{2}\mathbb{P}(Y \in A)$ is $|A \cap [0,1/2]|$. The second one is

$$|(A - 1/2) \cap [0,1/2]| = |A \cap [1/2,1]|$$

by invariance by translation of the Lebesgue measure. This gives $\mathbb{P}(\hat{X} \in A) = |A \cap [0,1]|$

Back to Exercise 2.15.

Solution to Exercise 2.16. Since $X_n \geq 0$, the $L^1(\Omega,\mathbb{P})$-norm of X_n is the integral

$$\int_\Omega X_n d\mathbb{P} = \int_{\{X_n=1\}} X_n d\mathbb{P} = \mathbb{P}(X_n = 1) = \frac{1}{n}.$$

Therefore $X_n \to 0$ in $L^1(\Omega,\mathbb{P})$. Let A be the event $\{X_n \to 0\}$. Using the $\varepsilon - n_0$ characterization of the convergence with $\varepsilon = k^{-1} < 1$, we obtain the usual description

$$A = \bigcap_{k\geq 1} \bigcup_{n\in\mathbb{N}} \bigcap_{p\geq n} \{|X_p| < k^{-1}\}.$$

Since X_n takes the values 1 or 0 only, this gives

$$A = \bigcup_{n\in\mathbb{N}} \bigcap_{p\geq n} \{X_p = 0\}.$$

Since $n \mapsto \bigcap_{p\geq n}\{X_p = 0\}$ is decreasing, the probability of A is

$$\mathbb{P}(A) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcap_{p\geq n}\{X_p = 0\}\right).$$

We introduce the intermediate sets $\bigcap_{m\geq p\geq n}\{X_p = 0\}$, which are decreasing with respect to m and the independence of the random variables (X_n) to obtain

$$\mathbb{P}(A) = \lim_{n\to+\infty}\lim_{m\to+\infty}\prod_{p=n}^{m}\mathbb{P}(X_p = 0).$$

Since $\mathbb{P}(X_p = 0) = 1 - \frac{1}{p}$, the product is divergent (use the log and compare to the harmonic series to justify this):

$$\lim_{m\to+\infty}\prod_{p=n}^{m}\mathbb{P}(X_p = 0) = 0,$$

for all n. Consequently, $\mathbb{P}(A) = 0$.

Note that our aim was initially to prove $\mathbb{P}(A) < 1$. We obtain much more: $\mathbb{P}(A) = 0$! According to the Kolmogorov's zero-one law, this was the only possible alternative.

Back to Exercise 2.16.

Solution to Exercise 2.17. We still have (2.12) by independence, where now A_i is a Borel subset of E_i. Also, the σ-algebra generated by the measurables rectangles $A = A_1 \times \cdots \times A_n$ is, by definition [Tao11, Section 1.7.4], the product σ-algebra $\mathcal{B}(E_1)\times\cdots\times\mathcal{B}(E_n)$. To conclude, we have to show that the product Borel σ-algebra coincides with the Borel σ-algebra $\mathcal{B}(E_1\times\cdots\times E_n)$ on the product. The proof is again similar to [Bil95, Example 18.1], replacing intervals by balls.

Back to Exercise 2.17.

Solution to Exercise 2.18. Let $X = X_1 + \cdots + X_n$. By iteration of Theorem 2.4 and Formula (2.13), we have

$$\int_{\mathbb{R}} h d\mu_X = \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} h(x_1+\cdots+x_n)d\mu(x_1)\cdots d\mu(x_n), \tag{B.2}$$

for all $h \in C_b(\mathbb{R})$, where $\mu = p\delta_1 + (1-p)\delta_0$. To compute the right-hand side of (B.2), we have to count the numbers of elements $(x_1,\ldots,x_n)\in\{0,1\}^n$ whose sum is a given number $k\in\{0,n\}$. Such elements have a contribution $p^k(1-p)^{n-k}$ in the right-hand side of (B.2). The question is therefore to evaluate the number of ways to pick up k elements (the x_i's with value 1) among n. There are $\binom{n}{k}$ of those elements, therefore

$$\int_{\mathbb{R}} h d\mu_X = \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}h(k).$$

The law of X is the Binomial law $\mathcal{B}(n,p)$.

Back to Exercise 2.18.

Solution to Exercise 2.20. We have

$$\mathbb{P}(X_n = [xn]) = \binom{n}{[xn]} p^{[xn]}(1-p)^{[xn]}.$$

Taking the ln of both sides gives

$$\ln[\mathbb{P}(X_n = [xn])] = \ln(n!) - \ln([xn]!) - \ln((n-[xn])!) + [xn]\ln(p) + (n-[xn])\ln(1-p).$$

We use the asymptotic development $\ln(n!) = n\ln(n) - n + o(n)$ to obtain, after simplifications,

$$\begin{aligned}\ln[\mathbb{P}(X_n = [xn])] =& n\left[-\frac{[xn]}{n}\ln\left(\frac{[xn]}{n}\right) - \left(1-\frac{[xn]}{n}\right)\ln\left(1-\frac{[xn]}{n}\right)\right.\\ &\left.+\frac{[xn]}{n}\ln(p) + \left(1-\frac{[xn]}{n}\right)\ln(1-p) + o(1)\right].\end{aligned}$$

This gives (2.17) with the rate function

$$H(x;p) = x\ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right).$$

We have $H(p;p) = 0$ and $H(x;p) > 0$ if $x \neq p$ by strict convexity of $-\ln$:

$$H(x;p) > -\ln\left[x\frac{p}{x} + (1-x)\frac{1-p}{1-x}\right] = \ln(1) = 0.$$

Back to Exercise 2.20.

Solution to Exercise 2.24. By induction, it is sufficient to consider the case $n=2$. Setting $Y_i = X_i - \mathbb{E}(X_i)$ if necessary, we can also assume $\mathbb{E}(X_i) = 0$. Consider first the case where $H = \mathbb{R}$. We have then $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2) = 0$ by (2.27). Developing the square $\mathbb{E}|X_1+X_2|^2$, we obtain the result. In the general case, let (e_n) be an orthonormal basis of H. Using Parseval's identity, we decompose, for $Z \in \{X_1, X_2, X_1+X_2\}$,

$$\mathbb{V}\mathrm{ar}(Z) = \mathbb{E}\|Z\|_H^2 = \mathbb{E}\sum_n |\langle Z, e_n\rangle_H|^2 = \sum_n \mathbb{V}\mathrm{ar}(\langle Z, e_n\rangle)$$

and use the real case to conclude.
Back to Exercise 2.24.

Solution to Exercise 2.27. For $\varphi \in C_b(\mathbb{R})$, we have

$$\mathbb{E}\varphi(X_n) = \varphi(0)\mathbb{P}(X_n = 0) + \varphi(1)\mathbb{P}(X_n = 1) \to \varphi(0).$$

To answer the second question, we may consider $\tilde{\Omega} = [0,1]$ with the σ-algebra $\tilde{\mathcal{F}}$ of Borel sets and the Lebesgue measure on $[0,1]$ as probability measure $\tilde{\mathbb{P}}$. Then we set $\tilde{X}_n = \mathbf{1}_{[0,n^{-1}]}$, $\tilde{X} = 0$. The identity of the laws is realized and $\tilde{X}_n \to 0$ $\tilde{\mathbb{P}}$-almost-surely. Note that the family $\{\tilde{X}_n; n \in \mathbb{N}^*\}$ is not independent.
Back to Exercise 2.27.

Solution to Exercise 2.30. Here is a proof using the characterization (2.32) of convergence in law. Let F be a closed subset of E. Let $\varepsilon > 0$ and $\delta > 0$. There exists an n_0 such that $\mathbb{P}(\|X_n - Y_n\|_E > \delta) < \varepsilon$ for all $n \geq n_0$. We have then $\mathbb{P}(Y_n \in F) < \varepsilon + \mathbb{P}(X_n \in \overline{F}^\delta)$, where $\overline{F}^\delta$ denotes the δ-neighbourhood of F:

$$\overline{F}^\delta = \{x \in E; d(x, F) \leq \delta\}, \quad d(x, F) = \min_{y \in F} \|x - y\|_E.$$

Since $\overline{F}^\delta$ is closed, we obtain

$$\limsup_{n \to +\infty} \mathbb{P}(Y_n \in F) \leq \varepsilon + \mu_X(\overline{F}^\delta).$$

Since $(\overline{F}^\delta) \downarrow F$ when $\delta \downarrow 0$ (because F is closed), we obtain $\limsup_{n \to +\infty} \mathbb{P}(Y_n \in F) \leq \varepsilon + \mu_X(F)$ at the limit $\delta \to 0$. Since ε is arbitrary, this gives the result. Note that we have repeated, more or less, the arguments of the proof of Proposition 2.8 and, indeed, we can use Proposition 2.8 and also the lines of the proof of Proposition 2.11 to write, for φ uniformly continuous and bounded, with a modulus of continuity denoted by ω_φ, that $|\mathbb{E}\varphi(Y_n) - \mathbb{E}\varphi(X)|$ is bounded by the sum of $|\mathbb{E}\varphi(X_n) - \mathbb{E}\varphi(X)|$ with

$$\begin{aligned}
&\mathbb{E}\left[|\varphi(X_n) - \varphi(Y_n)|\mathbf{1}_{\|X_n - Y_n\|_E > \delta}\right] + \mathbb{E}\left[|\varphi(X_n) - \varphi(Y_n)|\mathbf{1}_{\|X_n - Y_n\|_E \leq \delta}\right] \\
&\leq \|\varphi\|_{C_b(E)} \mathbb{P}(\|X_n - Y_n\|_E > \delta) + \omega_\varphi(\delta).
\end{aligned}$$

We choose first δ small, then n large to conclude.
Back to Exercise 2.30.

Solution to Exercise 2.31. For $\delta > 0$, set $\chi_\delta(s) = \delta^{-1}(\delta - s)^+$. Then $\varphi_\delta \colon (x, y) \mapsto \chi_\delta(\|x - y\|_E)$ is continuous on $E \times \mathbb{E}$ and we have

$$\mathbb{P}(\|X_n - X\|_E \leq \delta) = \mathbb{E}\mathbf{1}_{\|X_n - X\|_E \leq \delta} \geq \mathbb{E}\varphi_\delta(X_n, X).$$

the right-hand side is converging to $\varphi_\delta(Y, Y) = 1$ by hypothesis, which gives the result. Note that the result is true also when E is infinite dimensional.
Back to Exercise 2.31.

Solution to Exercise 2.33.

1. Let $K_n = [-n, n]$. We have $\mathbb{R} = \cup_{n \in \mathbb{N}} K_n$ (increasing union), hence $1 = \mu(\mathbb{R}) = \lim_{n \to +\infty} \mu(K_n)$. For all $\varepsilon > 0$, there exists an n such that $\mu(K_n) > 1 - \varepsilon$.

2. Same proof as in 1.

3. Note that tightness of $\{\mu\}$ is equivalent to the inner regularity of μ. If E is finite dimensional, then we can use Item 2. (E si equal to the increasing union of the close balls of centred at 0 with radius n, which are compact). In the infinite-dimensional case, we use the following characterization of compact sets in separable, complete,

metric (*i.e.* Polish) spaces: a set K is relatively compact if, and only if, for all $r>0$, it can be covered by a finite number of balls with radius r. Let $(r_n)\downarrow 0$ (sequence of radii) and let also $(\delta_n)\to 0$. By separability of E, there is a countable set $(B_n^k)_{k\in\mathbb{N}}$ of balls of radius r_n covering E (take the balls centred at each points x_k of a dense countable set). Therefore

$$1=\mu(E)=\lim_{k\to+\infty}\mu(D_n^k),\quad D_n^k:=\bigcup_{j=0}^k B_n^j,$$

and there exists k_n such that $\mu(D_n^{k_n})>1-\delta_n$. Let K be the closure of the set

$$A=\bigcap_{n\in\mathbb{N}} D_n^{k_n}.$$

Then K is compact and

$$\mu(K^c)\leq\mu(A^c)\leq\sum_{n\in\mathbb{N}}\mu(D_n^{k_n})<\sum_{n\in\mathbb{N}}\delta_n.$$

Taking δ_n such that $\sum_{n\in\mathbb{N}}\delta_n=\varepsilon$, we obtain the result.

4. By the Markov inequality, we have

$$\mathbb{P}(\|X_n\|_{H^1(\mathbb{T}^d)}>R)\leq\frac{1}{R}\mathbb{E}\|X_n\|_{H^1(\mathbb{T}^d)}\leq\frac{C}{R}.$$

This means, for $R>C$, $\mu_n(K_R)\geq 1-\frac{C}{R}$, where

$$K_R=\{u\in L^2(\mathbb{T}^d);\|u\|_{H^1(\mathbb{T}^d)\leq R}$$

is compact since the injection $H^1(\mathbb{T}^d)\hookrightarrow L^2(\mathbb{T}^d)$ is compact. If $\varepsilon>0$, we choose $R>C\varepsilon^{-1}$ to obtain $\mu_n(K_R)>1-\varepsilon$ for all n. This gives the result.

5. If we assume $\sup_n\mathbb{E}\|X_n\|_F<+\infty$ with $F\hookrightarrow E$ compact, we have the same result (with same proof): the family $\{\mu_{X_n};n\in\mathbb{N}\}$ is tight on E.

6. Reflected random walk. We can write $X_{n+1}-X_n=2\xi_n-1+\mathbf{1}_{X_n=0}$, where ξ_n is a Bernoulli of parameter p. By summing over n, this gives

$$X_n=2\left(S_n-\frac{n}{2}\right)+Y_n,\quad S_n=\sum_{k=0}^n\xi_k,\quad Y_n=\sum_{k=0}^{n-1}\mathbf{1}_{X_k=0}. \tag{B.3}$$

If $p>\frac{1}{2}$, we use the weak law of large number (Theorem 2.19): $\frac{S_n}{n}\to p$ in probability, therefore (since $Y_n\geq 0$), $X_n\to+\infty$ in probability, in the sense that, for all $R>0$, $\mathbb{P}(X_n\geq R)\to 1$. If $p=\frac{1}{2}$, we use the Central Limit Theorem:

$$\mathbb{P}\left(S_n-\frac{n}{2}\geq\sigma\sqrt{n}\right)\to c>0,$$

where $\sigma = \sqrt{p(1-p)}$ and $c = \mathbb{P}(Z \geq 1)$ for $Z \sim \mathcal{N}(0,1)$. Again we obtain $X_n \to +\infty$ in probability. Note that discarding the term Y_n in (B.3) because it is non-negative amounts to consider the non-reflected random walk. When $p < \frac{1}{2}$ now, we may compare (X_n) with the stationary solution (X_n^*). The stationary solution (invariant measure) is such that

$$\mathbb{P}(X_n^* = k) = \mathbb{P}(X_{n+1}^* = k) = p\mathbb{P}(X_n^* = k-1) + q\mathbb{P}(X_n^* = k+1), \tag{B.4}$$

if $k > 0$ (and $\mathbb{P}(X_n^* = 0) = \mathbb{P}(X_{n+1}^* = 0) = q(\mathbb{P}(X_n^* = 0) + \mathbb{P}(X_n^* = 1))$). We can solve (B.4) explicitly to find $\mathbb{P}(X_n^* = k) = (1-A)A^k$, $A := \frac{p}{q}$. Now we notice that $\mathbb{P}(X_0 = k) \leq C\mathbb{P}(X_n^* = k)$ for C large enough ($C = (1-A)^{-1}$ actually). This implies $\mathbb{P}(X_n = k) \leq C\mathbb{P}(X_n^* = k)$ for all n. It follows that (X_n) is tight.

Back to Exercise 2.33.

Solution to Exercise 2.34. In Example 2.10, one has to test (2.43) only for $A = \Omega$, which is satisfied then with $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$. In Example 2.11, we start from the fact that a $\mathcal{G}$-measurable function is of the form $\alpha\mathbf{1}_B + \beta\mathbf{1}_{B^c}$. Tested with $A = B$ (*resp.* B^c), the condition 2.43 gives $\alpha\mathbb{P}(B) = \mathbb{E}(\mathbf{1}_B X)$ (*resp.* $\beta\mathbb{P}(B^c) = \mathbb{E}(\mathbf{1}_{B^c} X)$. In Example 2.12, we have to prove that

$$\mathbb{E}\left[\mathbf{1}_A X\right] = \mathbb{E}\left[\mathbf{1}_A \mathbb{E}(X|\mathcal{G})\right],$$

for all $A \in \mathcal{H}$. But this is of course true since $\mathcal{H} \subset \mathcal{G}$. The identity (2.45) in Example 2.13 is a direct consequence of (2.43) with $A = \Omega$. At last, let us consider Example 2.14 By Theorem 2.1 we know that $\mathbb{E}(\Phi(X+Y)|\sigma(X))$, being $\sigma(X)$-measurable, is of the form $f(X)$. Any $\sigma(X)$-measurable set is of the form $A = X^{-1}(B)$ where B is a Borel subset of E. In that case, we have $\mathbf{1}_A = \mathbf{1}_B(X)$ and

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_B(X)\Phi(X,Y)) &= \int_{E\times E} \mathbf{1}_B(x)\Phi(x,y)d\mu_{(X,Y)}(x,y) \text{ by (2.30)},\\
&= \int_{E\times E} \mathbf{1}_B(x)\Phi(x,y)d(\mu_X \times \mu_Y)(x,y) \text{ by independence},\\
&= \int_E \mathbf{1}_B(x)\left[\int_E \Phi(x,y)d\mu_Y(y)\right] d\mu_X(x) \text{ by Fubini's Theorem},\\
&= \mathbb{E}[\mathbf{1}_B(X)f(X)], \quad f(x) := \mathbb{E}\Phi(x,Y) \text{ by (2.30) again}.
\end{aligned}$$

Back to Exercise 2.34.

Solution to Exercise 2.36. We apply first (2.46) in an obvious way to obtain

$$\mathbb{E}(X_{n+1}|\sigma(X_n)) = g_n(X_n), \quad g_n(x) := \mathbb{E}f(x, Y_{n+1}) = \int_F f(x,y)d\mu_{Y_{n+1}}(y).$$

Then we note that $\mathcal{F}_n = \sigma(Z)$, $Z = (Y_i)_{1,n}$ and that $X_{n+1} = \Phi(Z, Y_{n+1})$ to obtain, by (2.46),

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = h(Z), \quad h(z) = \mathbb{E}\Phi(z, Y_{n+1}) = \int_F \Phi(z,y)d\mu_{Y_{n+1}}(y).$$

Since $\Phi(z, y) = f(\dots f(f(x_0, z_1), z_2), \dots, z_n)$, we obtain $\Phi(Z, y) = f(X_n, y)$, which gives the result.
Back to Exercise 2.36.

Solution to Exercise 2.37. The statement of Theorem 2.6 is

$$\mu_n(A) \to \mu(A) \tag{B.5}$$

for all A of the form (a, b), where $\mu_n = \mu_{Z_n}$ and μ is the law of a $\mathcal{N}(0, 1)$ random variable. Let G be an open set in $\mathbb{R}$. Then G is the disjoint union of open intervals A_k, $k \in \mathbb{N}$. Let $\varepsilon > 0$. There exists $K \in \mathbb{N}$ such that $\mu(G) \leq \mu(G_K) + \varepsilon$, where G_K is the union of the intervals A_k over $k \in \{0, \dots, K\}$. Then (B.5) is true for $A = G_K$ and we deduce

$$\mu(G) \leq \lim_{n \to +\infty} \mu_n(G_K) + \varepsilon \leq \limsup_{n \to +\infty} \mu_n(G) + \varepsilon.$$

Since ε is arbitrary, this yields (2.33).
Back to Exercise 2.37.

Solution to Exercise 2.38. Let $Z = p_i \varphi(X_n + 1) - Y_i \varphi(X_n)$. We use (2.45) to obtain $\mathbb{E}(Z) = \mathbb{E}(\mathbb{E}(Z|\sigma(Y_i)))$. By (2.46) then, we have

$$\mathbb{E}(Z|\sigma(Y_i)) = f(Y_i), \quad f(y) := \mathbb{E}[p_i \varphi(X_n^{(i)} + y + 1) - y\varphi(X_n^{(i)} + y)]$$

and

$$\mathbb{E}(Z) = \mathbb{E}(f(Y_i)) = f(0)\mathbb{P}(Y_i = 0) + f(1)\mathbb{P}(Y_i = 1) = (1 - p_i)f(0) + p_i f(1).$$

This gives (2.55).
Back to Exercise 2.38.

Solution to Exercise 3.8. Let $A \in \mathcal{F}_{\text{cyl}}$ be a non-empty set, and let J be a countable subset of $[0, T]$, B an element of the cylindrical σ-algebra on E^J such that $A = \pi_J^{-1}(B)$. Let $t' \in [0, T] \setminus J$ and let x' be an arbitrary element of E. If $Y \in A$, then Y' defined by

$$Y' = Y \text{ on } [0, T] \setminus \{t'\}, \quad Y'(t') = x'$$

is also in A since the values Y_t for $t \in J$ are not affected by the modification of the value $Y_{t'}$. It is clear then that neither A_1 nor A_2 can be in $\mathcal{F}_{\text{cyl}}$. In the case of A_1, a contradiction is obtained by considering any $x' \neq 0$. In the case of A_2, a contradiction is obtained by considering any $x' \neq Y_{t'}$.
Back to Exercise 3.8.

Solution to Exercise 3.9. Let $\mathcal{B}$ denote the Borel σ-algebra on $C([0,T];E)$. Since each projection π_t from $C([0,T];E)$ onto E is continuous, the cylindrical sets are Borel sets, hence $\mathcal{F}_{\text{cts}} \subset \mathcal{B}$. To prove the converse inclusion, we consider an open ball $B(u,r)$ in $C([0,T];E)$. It can be described as

$$B(u,r) = \{v \in C([0,T];E); \forall n \in \mathbb{N}, |v(t_n) - u(t_n)| < r\},$$

where $(t_n)_{n\in\mathbb{N}}$ is a dense subset of $[0,T]$. This shows that $B(u,r) \in \mathcal{F}_\circ \cap C([0,T];E)$, where $\mathcal{F}_\circ$ is the σ-algebra generated by the sets introduced in (3.6). But we have shown that $\mathcal{F}_\circ = \mathcal{F}_{\text{cyl}}$. Therefore all open balls are in $\mathcal{F}_{\text{cts}}$, *i.e.* $\mathcal{B} \subset \mathcal{F}_{\text{cts}}$.

We have then $A_2 = C([0,T];E)$, the whole space, while A_1 is the singleton $\{0\}$, a closed set. Therefore $A_1, A_2 \in \mathcal{F}_{\text{cts}}$.

Back to Exercise 3.9.

Solution to Exercise 3.11.

1. In both cases $X_t = W_t$ or $X_t = N_t$, the law of X_t and the law of $X_{t+\sigma}$ are different ($\mathcal{N}(0,t)$ versus $\mathcal{N}(0,t+\sigma)$ in the case of the one-dimensional Wiener process; $\mathcal{P}(\lambda t)$ versus $\mathcal{P}(\lambda(t+\sigma))$ in the case of the Poisson process). Therefore (3.12) is not satisfied when $n=1$: the processes are not stationary.

2. Again, we consider (3.12) for $n=1$: it shows that the law of a stationary process is constant in time.

3. Note that we initialize the process with the measure μ_0 which is the invariant measure found in Exercise 2.8.

   (a) Let μ_n be the law of X_n. Let $\varphi \in C_b(\mathbb{R})$. For $n=1$, we have

   $$\begin{aligned}
   \langle \mu_1, \varphi\rangle = \mathbb{E}\varphi(X_1) &= \mathbb{E}\varphi(2^{-1}X_0 + Z_1)\\
   &= \iint_{\mathbb{R}^2} \varphi(2^{-1}x + z)\, d\mu_{(X_0,Z_1)}(x,z)\\
   &= \iint_{\mathbb{R}^2} \varphi(2^{-1}x + z)\, d\mu_{X_0}(x)\, d\mu_{Z_1}(z) \text{ by independence}\\
   &= \frac{1}{4} \times \frac{1}{2} \int_{-2}^{2} \left[\varphi(2^{-1}x - 1) + \varphi(2^{-1}x + 1)\right] dx\\
   &= \frac{1}{8}\left[2\int_{-2}^{0} \varphi(y)\,dy + 2\int_{0}^{2} \varphi(y)\,dy\right] = \langle \mu_0, \varphi\rangle.
   \end{aligned}$$

   By iteration of this computation, we obtain $\mu_n = \mu_0$ for all $n \in \mathbb{N}$.

   (b) Let $0 \le k_1 < \cdots < k_n \in \mathbb{N}$ and $l \in \mathbb{N}$. Let $B_1, \dots, B_n$ be some Borel subsets of $\mathbb{R}$. We want to show that

   $$\mathbb{P}(X_{k_1} \in B_1, \dots, X_{k_n} \in B_n) = \mathbb{P}(X_{k_1+l} \in B_1, \dots, X_{k_n+l} \in B_n). \tag{B.6}$$

We just showed the case $n = 1$. Assume $n = 2$. Intuitively, the identity (B.6) comes from the fact that the probability that, first X_{k_1} and X_{k_1+l} are in B_1 are the same, and, second, that, knowing that $X_m \in B_1$, the fact that $X_{m+p} \in B_2$ depends uniquely on the drawing of $(Z_{m+1}, \ldots, Z_p)$, which has the same law as $(Z_{m+1+l}, \ldots, Z_{p+l})$. Note also that Equation (3.13) may be replaced by the more general relation $X_{n+1} = f(X_n, Z_{n+1})$, the reasoning would be the same. The proof is the following one:

$$\begin{aligned}\mathbb{P}(X_{k_1+l} \in B_1, X_{k_2+l} \in B_2) &= \mathbb{E}(\mathbf{1}_{B_2}(X_{k_2+l})\mathbf{1}_{B_1}(X_{k_1+l})) \\ &= \mathbb{E}\left(\mathbb{E}\left[\mathbf{1}_{B_2}(X_{k_2+l})|\sigma(X_{k_1+l})\right]\mathbf{1}_{B_1}(X_{k_1+l})\right).\end{aligned} \tag{B.7}$$

For $\varphi \in \mathbb{R} \to \mathbb{R}$ measurable and bounded. By (2.46) applied to $X = X_{k_1+l}$, $Y = (Z_{k_1+l+1}, \ldots, Z_{k_2+l})$ we have

$$\mathbb{E}\left[\varphi(X_{k_2+l})|\sigma(X_{k_1+l})\right] = P_{k_2,k_1}\varphi(X_{k_1+l}),$$

where[11]

$$P_{k_1,k_2}\varphi(x) := \mathbb{E}\varphi(f(f(\ldots f(x, Z_{k_1+l+1}), \ldots,)Z_{k_2+l-1}, Z_{k_2+l}))).$$

Since $(Z_{k_1+l+1}, \ldots, Z_{k_2+l})$ has the same law as $(Z_{k_1+1}, \ldots, Z_{k_2})$, it makes sense to denote a dependence on k_1, k_2 solely in $P_{k_1,k_2}\varphi(x)$. Coming back to (B.7), we obtain

$$\begin{aligned}\mathbb{P}(X_{k_1+l} \in B_1, X_{k_2+l} \in B_2) =& \mathbb{E}\left(P_{k_2,k_1}\mathbf{1}_{B_2}(X_{k_1+l})\mathbf{1}_{B_1}(X_{k_1+l})\right) \\ &= \int_{\mathbb{R}} (P_{k_1,k_2}\mathbf{1}_{B_2}(x)\mathbf{1}_{B_1}(x)d\mu_0(x),\end{aligned}$$

since X_{k_1+l} has the law μ_0. This last expression is independent on l: this gives the desired result. The case of general n in (B.6) is obtained similarly by induction on n.

Back to Exercise 3.11.

Solution to Exercise 3.14. Consider the non-negative function

$$f = \sum_{n=0}^{\infty} \mathbf{1}_{A_n}.$$

By hypothesis, and thanks to Fubini's Theorem, we have

$$\mathbb{E}f = \int_\Omega f d\mathbb{P} = \sum_{n=0}^{\infty} \mathbb{P}(A_n) < +\infty.$$

Consequently, f is finite almost-surely. Equivalently, almost-surely, a finite number of the A_n's is realized.

Back to Exercise 3.14.

[11] the operator P_{k_1,k_2} is the transition operator, see Section 4

Solution to Exercise 3.15. Let $(\delta_k) \downarrow 0$. We apply the Borel-Cantelli lemma to the sets $A_n^k = \{\|X_n\|_E > \delta_k\}$: there exists $\Omega_k \subset \Omega$ of probability 1 such that all $\omega \in \Omega_k$ is in a finite number of A_n^k's: there exists $n_k(\omega)$ such that, for $n \geq n_k(\omega)$, $\omega \notin A_n^k$. Let $\tilde{\Omega} = \cap_k \Omega_k$. Then $\mathbb{P}(\tilde{\Omega}) = 1$ and if $\omega \in \tilde{\Omega}$ and $\varepsilon > 0$, then choosing k such that $\delta_k < \varepsilon$, we have $\|X_n(\omega)\|_E < \varepsilon$ for $n \geq n_k(\omega)$. This means $X_n(\omega) \to 0$.
Back to Exercise 3.15.

Solution to Exercise 3.16. Just take the modification furnished by the Kolmogorov Theorem and set

$$\zeta = C_{\sigma,p}^{1/p} \left[\int_0^T \int_0^T \frac{\|\tilde{X}(t') - \tilde{X}(s')\|_E^p}{|t'-s'|^{1+\sigma p}} ds' dt' \right]^{1/p},$$

where $\frac{1}{p} < \sigma < \frac{1+\delta}{p}$, $\sigma := \alpha + \frac{1}{p}$. We have $\mathbb{E}|\zeta|^p < +\infty$ (same computation as in (3.22)) and (3.28) thanks to (3.27).
Back to Exercise 3.16.

Solution to Exercise 3.17.

1. We will apply Lemma 2.12. By the Markov inequality, we have $\mathbb{P}(\|Y_n\|_E > \delta) \leq \delta^{-2}\mathbb{E}\|Y_n\|_E^2$ for $\delta > 0$, so (Y_n) is converging to 0 in probability. Therefore it is sufficient to show that $\eta_n := (a - a_n)X_n$ is converging to 0 in probability. For this we use the tightness of (X_n) (this is the "easy" part of the Prohorov theorem): given $\varepsilon > 0$, there exists a compact K such that $\mathbb{P}(X_n \in K) \geq 1 - \varepsilon$ for all n. There exists $R > 0$ such that $K \subset \bar{B}(0, R)$. It follows that

$$\mathbb{P}(\|\eta_n\|_E > \delta) \leq \varepsilon + \mathbb{P}(|a - a_n| > R^{-1}\delta).$$

   For n large enough, $\mathbb{P}(|a - a_n| > R^{-1}\delta) = 0$ (since (a_n) is deterministic here). This concludes the proof.

2. Clear with (2.11) (we use the generalization proved in Exercise 2.17).

Back to Exercise 3.17.

Solution to Exercise 4.3. See the correction of Exercise 3.11 for a proof in the time-discrete case. Since $(X_t)_{t\geq 0}$ is an homogeneous Markov process, (4.12) reads

$$\mu_{t_1,\dots,t_n} = (P_{t_n - t_{n-1}})^* \otimes \cdots (P_{t_2 - t_1})^* \otimes \mu_{t_1}.$$

If μ_{t_1} is independent on t_1, it is clear then that $\mu_{t_1,\dots,t_n} = \mu_{t_1+s,\dots,t_n+s}$ for all $s \geq 0$: $(X_t)_{t\geq 0}$ is stationary.
Back to Exercise 4.3.

Solution to Exercise 4.9. Assume Y is non trivial. Then, for $t > 0$, $\mathcal{F}_t^X = \sigma(Y)$ and thus $\mathcal{F}_{0+} = \sigma(Y)$ is distinct from $\{\emptyset, \Omega\} = \mathcal{F}_0$, although (X_t) has continuous trajectories.

Back to Exercise 4.9.

Solution to Exercise 4.11.

1. We have $P_n = P_1^n$ as a consequence of $P_0 = \mathrm{id}$, the relation $P_{n+1} = P_1 \circ P_n$ and recursion on n.

2. The collection of all sets of the form $E = B \cap D \cap \{N(t) = m\}$, $m \in \mathbb{N}$, $B \in \mathcal{F}_m^X$, $D \in \mathcal{F}_t^N$ form a π-system that generates $\mathcal{F}_t$. Therefore, by [Bil95, Theorem 3.3], it is sufficient to prove (4.28) for E as above. We have then

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_E \varphi(X_{n+N(t)})\right] =& \mathbb{E}\left[\mathbf{1}_{B\cap D\cap\{N(t)=m\}}\varphi(X_{n+m})\right] \\
=& \mathbb{P}(D \cap \{N(t) = m\})\mathbb{E}\left[\mathbf{1}_B \varphi(X_{n+m})\right] \text{ (independence)} \\
=& \mathbb{P}(D \cap \{N(t) = m\})\mathbb{E}\left[\mathbf{1}_B Q_n \varphi(X_m)\right] \\
=& \mathbb{E}\left[\mathbf{1}_{B\cap D\cap\{N(t)=m\}} Q_n \varphi(X_m)\right] \text{ (independence again)} \\
=& \mathbb{E}\left[\mathbf{1}_E Q^n \varphi(X_{N(t)})\right].
\end{aligned}$$

3. It follows from (4.28) that $\mathbb{E}\left[\varphi(X_{n+N(t)})|\mathcal{F}_t\right] = Q^n\varphi(X_{N(t)})$. We decompose

$$\mathbb{E}[\varphi(X_{N(t+s)})|\mathcal{F}_t] = \sum_{n=0}^{\infty} \mathbb{E}[\varphi(X_{N(t+s)})\mathbf{1}_{N(t+s)-N(t)=n}|\mathcal{F}_t]$$

and use independence to obtain

$$\begin{aligned}
\mathbb{E}[\varphi(X_{N(t+s)})|\mathcal{F}_t] =& \sum_{n=0}^{\infty} \mathbb{P}(N(t+s) - N(t) = n)\mathbb{E}[\varphi(X_{N(t)+n})|\mathcal{F}_t] \\
=& \sum_{n=0}^{\infty} e^{-s}\frac{s^n}{n!} Q_n \varphi(X_{N(t)}) \\
=& (\Pi_s \varphi)(X_{N(t)}),
\end{aligned}$$

with P_t defined by (4.30). It is clear that $\mathcal{L} = Q_1 - \mathrm{Id}$.

Back to Exercise 4.11.

Solution to Exercise 4.13.

1. If $\{\tau < t\} \in \mathcal{F}_t$ for all $t \geq 0$ then $\{\tau \leq t\} = \cap_{n\geq 1}\{\tau < t + n^{-1}\} \in \mathcal{F}_{t+}$. Conversely, if $\{\tau \leq t\} \in \mathcal{F}_{t+}$ for all $t \geq 0$, then $\{\tau < t\} = \cup_{n\geq 1}\{\tau \leq t(1 - n^{-1})\}$ is in $\mathcal{F}_t$.

2. By decomposing

$$\{\tau \wedge s \leq t\} = \Big(\{\tau \wedge s \leq t\} \bigcap \{\tau \leq s\}\Big) \bigcup \Big(\{\tau \wedge s \leq t\} \bigcap \{\tau > s\}\Big),$$

we obtain

$$\{\tau \wedge s \leq t\} = \{\tau \leq s \wedge t\} \bigcup \Big(\{s \leq t\} \bigcap \{\tau > s\}\Big) \in \mathcal{F}_{t\wedge s}.$$

If $A \in \mathcal{F}_{\tau\wedge s}$ then $A = A \cap \{tau \wedge s \leq s\} \in \mathcal{F}_s$. Therefore $\mathcal{F}_{\tau\wedge s} \subset \mathcal{F}_s$.

3. We have $\tau_A \leq t$ if, and only if, $\min_{s\in[0,t]} d(X_t, A) = 0$ since A is closed and $t \mapsto X_t$ is continuous. We deduce that

$$\{\tau_A \leq t\} = \bigcap_{n\geq 1} \bigcup_{s\in\mathbb{Q}\cap[0,t]} \{d(X_t, A) < n^{-1}\} \in \mathcal{F}_t.$$

This shows that $\tau \wedge s$ is a stopping time and that it is a $\mathcal{F}_s$-measurable random variable.

4. We have

$$\{\tau_A < t\} = \bigcup_{s<t} \{X_s \in A\} = \bigcup_{s\in\mathbb{Q}, s<t} \{X_s \in A\}. \tag{B.8}$$

The first equality in (B.8) is clear: $\tau_A \geq t$ means that X_s does not meet A for all $s < t$. The second equality in (B.8) uses the fact that A is open and (X_t) is right-continuous: if $X_s \in A$, then $X_s \in B$ where B is an open ball contained in A. Since $X_\sigma \to X_s$ when $\sigma \downarrow s$, there exists $\sigma \in \mathbb{Q} \cap (s,t)$ such that $X_sigma \in B \subset A$. By (B.8), we have $\{\tau_A < t\} \in \mathcal{F}_{t+}$. By Question 1, we deduce that τ_A is an $(\mathcal{F}_{t+})$-stopping time.

5. We use the notation (and result) of Remark 4.6. Let $Y(t) = X(t \wedge \tau)$. We have

$$Y(t) = \sum_{i=1}^{m} X(t \wedge \tau) \mathbf{1}_{\tau = t_i} = \sum_{i=1}^{m} X(t \wedge t_i) \mathbf{1}_{\tau = t_i},$$

which shows that $Y(t)$ is a random variable and that, for $B \in \mathcal{B}(E)$,

$$\{Y(t) \in B\} = \bigcup_{i=1}^{m} \{X(t \wedge t_i) \in B\} \cap \{\tau = t_i\}.$$

Back to Exercise 4.13.

Solution to Exercise 5.2. For $0 \le s \le t$, we have $X_t = X_s + \delta$, where δ is independent from $\mathcal{F}_s$, *i.e.* $\mathbb{E}[\delta|\mathcal{F}_s] = \mathbb{E}[\delta] = 0$. Consequently, $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$: $(X_t)_{t\ge 0}$ is a martingale for $(\mathcal{F}_t)_{t\ge 0}$. We also have

$$X_t^2 = (X_s + \delta)^2 = X_s^2 + 2\delta X_s + \delta^2.$$

Taking the conditional expectancy with respect to $\mathcal{F}_s$ and using independence again, we obtain

$$\mathbb{E}[X_t^2|\mathcal{F}_s] = X_s^2 + \mathbb{E}[\delta^2]. \tag{B.9}$$

Taking expectation in (B.9) gives $\mathbb{E}[X_t^2] = \mathbb{E}[X_s^2] + \mathbb{E}[\delta^2]$, hence $\mathbb{E}[X_t^2] \ge \mathbb{E}[X_s^2]$, but also (replacing $\mathbb{E}[\delta^2]$ by $\mathbb{E}[X_t^2] - \mathbb{E}[X_s^2]$ in (B.9))

$$\mathbb{E}[X_t^2|\mathcal{F}_s] - \mathbb{E}[X_t^2] = X_s^2 - \mathbb{E}[X_s^2].$$

This shows that $(X_t^2 - \mathbb{E}[X_t^2])_{t\ge 0}$ is a martingale.

Back to Exercise 5.2.

Solution to Exercise 5.3.

1. The function φ^* is the sup of affine functions. It is convex on $\mathbb{R}$ consequently, and thus continuous. This gives (a). Since φ^* is continuous, any countable dense subset D of $\mathbb{R}$ will do: we obtain (b). For $p \in D$, we have $\mathbb{E}[pX - \varphi^*(p)|\mathcal{G}] = p\mathbb{E}[X|\mathcal{G}] - \varphi^*(p)$ a.s. That $X \mapsto \mathbb{E}[X|\mathcal{G}]$ is monotone non-decreasing is clear from the definition (2.43) since $\mathbf{1}_A$ is non-negative. Therefore $\mathbb{E}[\varphi(X)|\mathcal{G}] \ge p\mathbb{E}[X|\mathcal{G}] - \varphi^*(p)$ a.s. Taking the sup on $p \in D$, we obtain the result (c). That $(\varphi(X_t))$ is a sub-martingale is then a direct consequence of (5.1).

2. Note first that E is separable since E^* is separable, [Bre11, Theorem 3.26]. Let D be a countable dense subset of the closed unit ball $\bar{B}^*$ of E^*. We have

$$\|x\|_E = \sup_{p\in\bar{B}^*} \langle p, x\rangle = \sup_{p\in D} \langle p, x\rangle. \tag{B.10}$$

   The first identity in (B.10) is [Bre11, Corollary 1.3]. Note that (B.10) is the identity $\varphi = \varphi^{**}$ for $\varphi(x) = \|x\|_E$. Indeed, it is easy to compute

$$\varphi^*(p) := \sup_{x\in E}[\langle p, x\rangle - \varphi(x)] = \begin{cases} 0 & \text{if } \|p\|_{E^*} \le 1, \\ +\infty & \text{if } \|p\|_{E^*} > 1. \end{cases}$$

   Once we have (B.10), the proof follows as in 1.

3. Consequence of (5.2).

Back to Exercise 5.3.

Solution to Exercise 5.4. Assume that Y can be decomposed as $Y = M + A$ as required. We have then $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ by the martingale property of M. Replacing M by $Y - A$, we obtain $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] - A_{n+1} = Y_n - A_n$ since A_{n+1} is $\mathcal{F}_n$-measurable. It is sufficient therefore to define (A_n) recursively by the formula

$$A_0 = 0, \quad A_{n+1} = A_n + \mathbb{E}[Y_{n+1}|\mathcal{F}_n] - Y_n,$$

to obtain the desired decomposition. We see that A is non-decreasing precisely because Y is a submartingale. The uniqueness of the decomposition comes from the fact that a predictable martingale is constant.

Back to Exercise 5.4.

Solution to Exercise 5.5. The answer is given by Exercise 5.2, where we have shown that $\langle M, M\rangle_t = \mathbb{E}|M_t|^2$ if (M_t) is a continuous martingale with independent increments. For the one-dimensional Wiener process (B_t), we obtain $\langle B, B\rangle_t = t$.

Back to Exercise 5.5.

Solution to Exercise 5.6. Let $\varepsilon > 0$ and let δ be the modulus of uniform continuity of $t \mapsto X(t)$ associated to ε. Using the identity (*cf.* (5.32))

$$\mathbb{E}\sum_k |\zeta(t_k)|^2 = |X(t_K)|^2 - |X(0)|^2 \leq M^2, \tag{B.11}$$

we have

$$\mathbb{E}|D|^2 \leq M^2\varepsilon^2 + \sum_k \mathbf{1}_{|\zeta(t_k)|\geq\varepsilon}|\zeta(t_k)|^4. \tag{B.12}$$

Let $N > 0$. Let γ_N denote the stopping time

$$\gamma_N = \min\left\{t_k; \sum_{i=1}^{k} \mathbf{1}_{|\zeta(t_i)|\geq\varepsilon} = N\right\} \cup \{t_K\}$$

and let κ_N be such that $\gamma_N = t_{\kappa_N}$. We have (same proof as (5.40))

$$\mathbb{E}|D|^2 \leq M^2\varepsilon^2 + M^2\mathbb{P}(\gamma_N < t_K) + M^2N\mathbb{P}(\delta < |\sigma|),$$

Taking N large, then $|\sigma|$ small gives the result.

Back to Exercise 5.6.

Solution to Exercise 5.7. We consider the quantity $\mathbb{E}\left[|X_{t_{i+1}} - X_{t_i}|^2|\mathcal{F}_{t_i}\right]$ for $t_{i+1}-t_i$ small. Let us discuss the occurrence of jumps between t_i and t_{i+1}. Denote by B_k the event corresponding to the occurrence of exactly k jumps of the Poisson process between t_i and t_{i+1} and set $B_2^+ = \cup_{k\geq 2}B_k$. We have

$$\begin{aligned}\mathbb{E}\left[|X_{t_{i+1}} - X_{t_i}|^2|\mathcal{F}_{t_i}\right] = \mathbb{E}\left[|X_{t_{i+1}} - X_{t_i}|^2\mathbf{1}_{B_0}|\mathcal{F}_{t_i}\right] + \mathbb{E}\left[|X_{t_{i+1}} - X_{t_i}|^2\mathbf{1}_{B_1}|\mathcal{F}_{t_i}\right]\\ + \mathbb{E}\left[|X_{t_{i+1}} - X_{t_i}|^2\mathbf{1}_{B_2^+}|\mathcal{F}_{t_i}\right].\end{aligned}$$

Since $|X_{t_{i+1}} - X_{t_i}|^2 \mathbf{1}_{B_0} = 0$ and $\mathbb{P}(B_2^+) = \mathcal{O}(|t_{i+1} - t_i|^2)$, the only term that matters is the one corresponding to the occurrence of exactly one jump t_i and t_{i+1}. By independence, and since $\mathbb{P}(B_1) = t_{i+1} - t_i$, we have

$$\mathbb{E}\left[|X_{t_{i+1}} - X_{t_i}|^2 \mathbf{1}_{B_1} | \mathcal{F}_{t_i}\right] = (\langle \varphi, \nu \rangle - \varphi(X_{t_i}))(t_{i+1} - t_i)$$

where φ is the function $x \mapsto x^2$, *i.e.*

$$\mathbb{E}\left[|X_{t_{i+1}} - X_{t_i}|^2 \mathbf{1}_{B_1} | \mathcal{F}_{t_i}\right] = (\mathcal{L}\varphi)(X_{t_i})(t_{i+1} - t_i).$$

Therefore, we infer the limit

$$\int_0^t (\mathcal{L}\varphi)(X_s) ds$$

for $\bar{V}_\sigma^{(2)}(t)$. Compare with the statement of Theorem 5.8 (Dynkin's formula).

Back to Exercise 5.7.

Solution to Exercise 7.1.

1. The elementary process
$$g_\sigma := \sum_{k=0}^{n-1} g(t_k) \mathbf{1}_{(t_k, t_{k+1}]} \tag{B.13}$$
is converging to g in $L^2(\Omega \times [0,T])$. Indeed,
$$\int_0^T \mathbb{E}|g(t) - g_\sigma(t)|^2 dt = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \mathbb{E}|g(t_k) - g(t)|^2 dt \leq T\omega(g; |\sigma|),$$
where the modulus of continuity
$$\omega(g; \delta) = \sup\left\{\mathbb{E}|g(t) - g(s)|^2; s, t \in [0,T], |s - t| < \delta\right\}$$
tends to 0 when $\delta \to 0$.

2. We will show that $g \in C([0,T]; L^2(\Omega))$ and apply Question 1. Let $\varepsilon > 0$, let δ be a (random) modulus of continuity associated to ε. We have, for $t \in (0,T)$, and $|s|$ smaller than $\min(t, T-t)$,
$$\begin{aligned}\mathbb{E}|g(t+s) - g(t)|^2 =& \mathbb{E}\left[\mathbf{1}_{s<\delta}|g(t+s) - g(t)|^2\right] + \mathbb{E}\left[\mathbf{1}_{s\geq\delta}|g(t+s) - g(t)|^2\right] \\ &\leq \varepsilon^2 + (\mathbb{E}\left[|g(t+s) - g(t)|^q\right])^{2/q} \mathbb{P}(s \geq \delta)^{\frac{q-2}{2}} \\ &\leq \varepsilon^2 + (2C)^{2/q} \mathbb{P}(s \geq \delta)^{\frac{q-2}{2}}, \quad C = \sup_{t\in[0,T]} \mathbb{E}|g(t)|^q.\end{aligned}$$
This gives the result since $\mathbb{P}(s \geq \delta) \to 0$ when $s \to 0$. We use the same reasoning when $t = 0$ or $t = T$.

3. Since $C([0,T])$ is dense in $L^2(0,T)$, we may assume that g is continuous. Then use Question 1. Since
$$\sum_{k=0}^{n-1} g(t_k)(\beta(t_{k+1}) - \beta(t_k))$$
is a linear combination of independent Gaussian random variables, it is a Gaussian random variables and the limit in $L^2(\Omega)$ of Gaussian random variables is a Gaussian random variable. The value of σ^2 follows from the Itô isometry.

Back to Exercise 7.1.

Solution to Exercise 8.2.

1. We have
$$X_t = e^{-t}x + \sqrt{2}\int_0^t e^{-(t-s)}dB_s.$$
This is a Gaussian random variable with mean $e^{-t}x$ and (by independence of the components of the d-dimensinoal Wiener process) diagonal covariance $\sigma^2\mathrm{Id}$, where, using Itô's isometry, we have
$$\sigma^2 = \mathbb{E}\left|\sqrt{2}\int_0^t e^{-(t-s)}dB_s\right|^2 = 2\int_0^2 e^{-2(t-s)}ds = 1 - e^{-2t}.$$
It follows that $X_t \to \mathcal{N}(0, \mathrm{Id})$ in law when $[t \to +\infty]$.

2. Itô's Formula gives, for $Y_t = \ln|X_t|$,
$$dY_t = -\frac{\sigma^2}{2}dt + \sigma dB_t.$$
We obtain $Y_t = Y_0 - \frac{\sigma^2}{2}t + \sigma B_t$ and
$$X_t = x\exp\left(-\frac{\sigma^2}{2}t + \sigma B_t\right).$$

Back to Exercise 8.2.

C The end

References

[Bau14] Fabrice Baudoin, *Diffusion processes and stochastic calculus*, EMS Textbooks in Mathematics, European Mathematical Society (EMS), Zürich, 2014.

[BDG72] D. L. Burkholder, B. J. Davis, and R. F. Gundy, *Integral inequalities for convex functions of operators on martingales*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, Univ. California Press, Berkeley, Calif., 1972, pp. 223–240.

[BE83] A. D. Barbour and G. K. Eagleson, *Poisson approximation for some statistics based on exchangeable trials*, Advances in Applied Probability **15** (1983), no. 3, 585–600. MR 706618

[Bil95] P. Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1995, A Wiley-Interscience Publication.

[Bil99] ———, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication.

[Bre11] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011.

[BSV12] M. Beiglböck, W. Schachermayer, and B. Veliyev, *A short proof of the Doob-Meyer theorem*, Stochastic Process. Appl. **122** (2012), no. 4, 1204–1209.

[Che75] Louis H. Y. Chen, *Poisson approximation for dependent trials*, Ann. Probability **3** (1975), no. 3, 534–545. MR 0428387

[CW90] K. L. Chung and R. J. Williams, *Introduction to stochastic integration*, second ed., Probability and its Applications, Birkhäuser Boston, Inc., Boston, MA, 1990.

[DPZ92] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications, vol. 44, Cambridge University Press, Cambridge, 1992.

[Dur84] R. Durrett, *Brownian motion and martingales in analysis*, Wadsworth Mathematics Series, Wadsworth International Group, Belmont, CA, 1984.

[EK86] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986, Characterization and convergence.

[Eva10] Lawrence C. Evans, *Partial differential equations*, second ed., Graduate Studies in Mathematics, vol. 19, American Mathematical Society, Providence, RI, 2010. MR 2597943

[FH14] Peter K. Friz and Martin Hairer, *A course on rough paths*, Universitext, Springer, Cham, 2014, With an introduction to regularity structures. MR 3289027

[JS03] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003.

[MR14] C. Marinelli and M. Röckner, *On maximal inequalities for purely discontinuous martingales in infinite dimensions*, Séminaire de Probabilités XLVI, Lecture Notes in Math., vol. 2123, Springer, Cham, 2014, pp. 293–315.

[Paz83] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences, vol. 44, Springer New York, New York, NY, 1983.

[Pri99] E. Priola, *On a class of Markov type semigroups in spaces of uniformly continuous and bounded functions*, Studia Math. **136** (1999), no. 3, 271–295.

[Pro05] Philip E. Protter, *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer-Verlag, Berlin, 2005.

[PSV77] G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan, *Martingale approach to some limit theorems*, Papers from the Duke Turbulence Conference (Duke Univ., Durham, N.C., 1976), Paper No. 6, Duke Univ., Durham, N.C., 1977, pp. ii+120 pp. Duke Univ. Math. Ser., Vol. III.

[RY99] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999.

[Ste72] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory, Univ. California Press, Berkeley, Calif., 1972, pp. 583–602. MR 0402873

[Tao11] Terence Tao, *An introduction to measure theory*, Graduate Studies in Mathematics, vol. 126, American Mathematical Society, Providence, RI, 2011. MR 2827917

[Yos80] K. Yosida, *Functional analysis*, sixth ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 123, Springer-Verlag, Berlin-New York, 1980.